Leaving the Nest: Simulated Cardell Errors for Flexible Dependence in Panel Discrete Choice*

Aaron P. Kaye
Boston University

Zach Y. Brown
University of Michigan & NBER

March 26, 2026

[For Conference Review, Please do not Circulate]

Abstract

We develop a modular Cardell variance-components framework for panel discrete-choice demand. By specifying which latent components are shared across products and time, the framework accommodates rich dependence in unobserved utility while preserving features that make logit models tractable. The representation extends Cardell beyond nested logit and unifies nesting structures, persistent match effects, and autoregressive shocks. We develop one specification in detail: an AR-persistent model in which serial correlation in unobserved preferences arises from stable consumer-product match quality and transitory shocks that decay over purchase occasions. We estimate the model with a pairwise composite likelihood that preserves the exact single-period logit likelihood and requires simulation only for the dependence correction. Monte Carlo experiments show accurate recovery of parameters and elasticities. We illustrate the approach by embedding AR-persistent errors in a standard empirical IO demand framework with endogenous prices, using linked household-panel and scanner data on over-the-counter cough remedies. Relative to mixed logit benchmarks with and without lagged-choice controls, the AR-persistent model better matches repeat purchases, returns after stockouts, and long-horizon out-of-sample prediction. In our application, models that ignore serial correlation in unobserved preferences overstate price sensitivity and welfare losses in product-removal counterfactuals.

JEL: L0, C51, C53
Keywords: discrete choice, panel data, persistent preferences, serial correlation

*Researchers' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. We thank Camila Bratten for research assistance.

1 Introduction

Discrete-choice demand models are central for estimating price sensitivities, characterizing substitution patterns, and evaluating counterfactuals involving prices, product availability, and entry. Many applications now rely on long household or consumer panels in which the same decision-maker makes repeated choices among a large set of differentiated products. Standard logit-based models deliver tractability, but they sharply restrict the dependence patterns that can be introduced into the unobserved component of utility. Nested logit accommodates correlation among close substitutes, and random-coefficients logit accommodates heterogeneity in tastes for observed characteristics, but both typically maintain that remaining shocks are independent over time. In long panels, this restriction is often at odds with observed persistence in choices and can lead to counterfactual predictions that are sensitive to how persistence is attributed to state dependence, time-invariant preferences, or measurement error.

We begin by developing a general framework for panel discrete-choice demand in which dependence across products and over time is introduced through layered Cardell (log-positive stable) components, while conditional choice probabilities remain multinomial logit. The construction exploits the identity in Cardell (1997) that a Cardell shock plus a suitably scaled type-I extreme-value shock is again type-I extreme value. This property has primarily been used in the context of the nested logit model. In our setting, this convolution property allows us to separate the component of utility that carries correlation and persistence from the component that delivers logit tractability: the former is represented by a low-dimensional set of Cardell components that are simulated directly, while the latter is integrated out analytically. The framework extends the role of the Cardell distribution beyond its standard use in nested logit and provides a modular representation of dependence in panel demand.

We then show that the framework nests a range of familiar and useful models and makes these cases practical for estimation. Single- and multi-level nested logit, cross-classified within-period correlation structures, persistent household-product match effects, and finite-memory autoregressive shocks all arise as special cases of the same representation. A practical obstacle has been that Cardell draws are difficult to simulate directly. Building on the log-stable representation in recent work by Galichon (2022), we use a simulation approach for Cardell draws based on fixed underlying randomness, which yields smooth objective functions in the dependence parameters and is well suited to standard simulation methods. Because each purchase occasion is standard logit conditional

on these draws, the same computational approach applies across special cases and richer layered specifications.

To demonstrate the empirical relevance of the framework, we focus on a parsimonious AR-persistent specification and an estimation strategy suited to long panels. The specification combines a persistent match component with autoregressive innovations whose influence decays across purchase occasions, allowing the data to distinguish stable household-product affinity from transitory shocks. For shorter panels, the model can be estimated by simulated maximum likelihood that draws only the Cardell components. For longer panels, we use a pairwise composite-likelihood estimator that preserves the exact single-period logit contribution and identifies the dependence parameters from low-dimensional intertemporal terms. The key tractability gain is therefore not only that the remaining integrals are lower dimensional, but also that a large share of the criterion can still be evaluated analytically. Relative to Gaussian specifications, the computational burden is shifted from a full-panel integral to a collection of low-dimensional pairwise terms, which makes the AR-persistent specification practical in applications with large choice sets and long purchase histories.[1]

We illustrate the practical value of this approach in an application to the over-the-counter (OTC) cough-remedies category using linked NielsenIQ scanner and household-panel data. In this setting, consumer choices exhibit substantial persistence, with households repeating their previous brand-formulation choice on more than half of purchase occasions. Using the approach of Grieco et al. (2025), we also show that the framework readily accommodates instruments and the joint use of aggregate shares and micro data. We compare the AR-persistent specification to standard random-coefficients logit benchmarks, including models with reduced-form lagged-choice state dependence. We find that standard i.i.d. error models struggle to jointly match observed repeat-purchasing and return-after-stockout patterns, and consequently overstate heterogeneity in price sensitivity in order to rationalize unmodeled persistence. By contrast, the AR-persistent model more accurately captures both short- and long-run persistence.

These differences matter for out-of-sample prediction and welfare analysis. The AR-persistent model delivers substantially higher predictive accuracy, especially over longer forecast horizons. In a partial-equilibrium counterfactual that removes the market-leading product, models that rely on lagged-choice state dependence imply large consumer-surplus losses because the forced switch mechanically lowers utility. By contrast, when persistence is attributed to stable match quality and

[1] A small related literature uses pairwise likelihood in probit panel models. See, for example, Tuzcuoglu (2023) and the overview in Bhat (2014). Relative to the case of Gaussian errors, the Cardell structure preserves exact logit single-period terms, so the pairwise term is especially tractable.

gradually decaying shocks, the implied welfare losses are substantially smaller and concentrated among the households most attached to the removed product. Taken together, the results underscore that flexibly modeling the unobserved error structure is important for reliable welfare and policy evaluation in panel demand settings.

Related Literature

The Cardell distribution is closely associated with the nested logit model. Cardell (1997) identified the specific distribution required for the "within-nest" error in nested logit, and subsequent work has often referred to the distribution as "Cardell." The Cardell distribution entered applied work mostly through its existence result rather than through direct simulation. Its CDF does not admit a closed-form expression, which historically made simulation of the Cardell distribution computationally inconvenient in practice. As a result, applied work typically exploited Cardell's construction indirectly via analytic GEV probabilities (e.g., Berry, 1994).[2] A related literature uses the Cardell distribution in models with costly information acquisition to derive a tractable estimation equation (Brown and Jeon, 2024; Boehm, 2025; Brown et al., 2026). A literature in matching theory also utilizes the Cardell distribution for tractability (Galichon and Salanié, 2022).

Recent advances by Galichon (2022) show that the Cardell distribution is the scaled logarithm of a positive stable distribution. This new insight allows for computationally fast approaches to generate simulated draws from the Cardell distribution. In prior work, Kaye (2024) uses simulated draws from the Cardell distribution in a search model to decompose match quality into a visible component (known pre-search) and a hidden component (revealed through costly search). The ability to simulate draws from the Cardell distribution makes it practical to use Cardell components in specifications that still preserve many of the computational benefits of working with EV1 distributions, but go well beyond standard nesting, which is the focus of this paper.

Our AR-persistent specification contributes to the literature seeking to disentangle the overlapping sources of persistence in panel discrete choice data. In many demand settings, repeated purchasing can reflect stable household–product affinity, reduced-form state dependence or switching costs tied to past choices, and serially correlated taste shocks that decay gradually. Because these mechanisms imply different counterfactual responses a broad applied literature has developed

[2]The Generalized Extreme Value (GEV) family of discrete-choice models delivers closed-form choice probabilities for static dependence across alternatives within a choice occasion. Our framework instead leverages simulation to extend this logic to more flexible dependence structures, including dependence over time in panels, while keeping choice probabilities in a logit form conditional on simulated components.

methods aimed at separating them, often using reduced-form state dependence or distributed-lag formulations (Dubé et al., 2010; Seetharaman, 2004).

Despite the importance of accounting for these issues, the literature has emphasized the computational barriers to incorporating serially correlated errors and persistent heterogeneity when estimating discrete choice models using long panels (Keane, 2015). The likelihood depends on the joint probability of an individual's full choice history, so allowing for serial correlation or persistent heterogeneity typically entails a high-dimensional integral that grows with the length of the panel—especially in Gaussian (probit) formulations (Keane, 2015; Berry and Haile, 2021). Early parametric discrete choice work tackled the high-dimensional sequence-likelihood and initial-conditions burdens of panel choice models using simulation-based estimators and related conditional approaches and often necessitated significant simplifications (e.g., Keane, 1997; Chintagunta et al., 2001; Cantillo et al., 2007).[3] Conditional or matching estimators provide another route under stronger support conditions on covariates and choice sets (Honore and Kyriazidou, 2000). More recent semiparametric work avoids specifying the error distribution and instead exploits revealed-preference restrictions to difference out fixed effects (Shi et al., 2018; Pakes and Porter, 2024), and extends these ideas to bound state dependence and switching costs in rich panel environments (Pakes et al., 2025). We provide a complementary, highly tractable parametric solution that preserves exact single-period logit structure while flexibly capturing both persistent match quality and gradually decaying shocks, enabling household-specific prediction in long panels.

The remainder of the paper proceeds as follows. Section 2 presents the general framework and the simulation approach. Section 3 develops benchmark special cases and the AR-persistent specification. Section 4 discusses estimation and reports Monte Carlo evidence. Section 5 presents the empirical application. Section 6 concludes.

2 A Generalized Cardell Error Structure

2.1 Panel choice environment

We study demand in a panel discrete-choice setting in which consumers $i = 1, \ldots, N$ make repeated choices over purchase occasions $t = 1, \ldots, T$ from a set of products $j \in \{0, 1, \ldots, J\}$, where $j = 0$

[3]The literature has also proposed Bayesian estimation to allow for AR errors in discrete choice models (Lachaab et al., 2006), as well as GMM or simulated method of moments approaches that match low-dimensional intertemporal correlations rather than the full sequence likelihood (Keane, 1994; Bertschek and Lechner, 1998). Bayesian methods have also been used to address related issues with serially correlated errors or persistent unobservables in dynamic discrete choice models (Norets, 2009; Blevins, 2016).

denotes the outside option. Let $a_{jt} \in \{0,1\}$ indicate whether product j is available at t; we take the outside option as always available $a_{0t} \equiv 1$. Consumer i's utility from choosing j at occasion t is

$$u_{ijt} \;=\; \delta_{jt} \;+\; \mu_{ijt} \;+\; \epsilon_{ijt}, \qquad j = 0, 1, \ldots, J,\; t = 1, \ldots, T, \tag{2.1}$$

where δ_{jt} is mean utility common across consumers, μ_{ijt} captures consumer-specific taste heterogeneity, and ϵ_{ijt} collects unobserved determinants of utility.

For most of the framework we develop below, we simplify by abstracting from taste heterogeneity and setting

$$\mu_{ijt} \equiv 0.$$

A convenient benchmark is then

$$\delta_{jt} \;=\; x_{jt}'\beta \;-\; \alpha p_{jt} \;+\; \xi_j, \tag{2.2}$$

where p_{jt} is price, x_{jt} are observed characteristics, and ξ_j is a product fixed effect (or unobserved time-invariant quality). The results below do not rely on $\mu_{ijt} \equiv 0$; we reintroduce μ_{ijt} in the empirical application to accommodate household heterogeneity through random coefficients as in Berry et al. (1995).

The central object is the joint distribution of ϵ_{ijt} across alternatives and time. Our goal is to allow flexible correlation patterns in ϵ_{ijt}, including within-nest correlation, match-specific persistence, and dynamic dependence over t—while retaining logit choice probabilities conditional on a low-dimensional set of latent variables.

2.2 The Cardell distribution

Let EV1 denote the standard type-I extreme value (Gumbel) distribution with cumulative distribution function $F(x) = \exp\{-\exp(-x)\}$. If $\varepsilon \sim \text{EV1}$, then $\lambda\varepsilon$ is type-I extreme value with scale parameter λ.

Definition 1 (Cardell distribution)**.** Fix $\lambda \in (0, 1]$. A random variable ν is said to follow a Cardell distribution with parameter λ, written $\nu \sim C(\lambda)$, if for any independent $\varepsilon \sim \text{EV1}$,

$$\nu + \lambda\varepsilon \sim \text{EV1}. \tag{2.3}$$

Definition 1 defines the Cardell distribution through the conjugacy relation that underlies our construction. The Cardell component ν exactly fills in the dispersion missing from the scaled extreme-value term $\lambda\varepsilon$, so that the sum has a standard EV1 marginal. In the limiting case $\lambda = 1$, the only solution is $\nu \equiv 0$. Smaller values of λ correspond to a larger Cardell component and, in our applications, greater scope for dependence when ν is shared across alternatives or over time. Thus, shared Cardell components carry correlation and persistence, while the remaining i.i.d. EV1 term preserves conditional logit tractability.

A second fact used repeatedly below is the composition rule

$$\nu_1 \sim C(\lambda_1), \qquad \nu_2 \sim C(\lambda_2), \qquad \nu_1 \perp \nu_2 \quad \Longrightarrow \quad \nu_1 + \lambda_1\nu_2 \sim C(\lambda_1\lambda_2). \tag{2.4}$$

To verify (2.4), let $\varepsilon \sim \text{EV1}$ be independent of (ν_1, ν_2). Then $\nu_2 + \lambda_2\varepsilon \sim \text{EV1}$, so by Definition 1,

$$\nu_1 + \lambda_1\nu_2 + \lambda_1\lambda_2\varepsilon = \nu_1 + \lambda_1(\nu_2 + \lambda_2\varepsilon) \sim \text{EV1},$$

which is precisely the defining property of $C(\lambda_1\lambda_2)$. Equation (2.4) is the device behind our layered error structures: sums of independent Cardell primitives can be collapsed into a single Cardell term, while (2.3) guarantees that adding the appropriately scaled EV1 remainder restores a standard EV1 marginal.

While $C(\lambda)$ does not admit a closed-form CDF, historically making it difficult to simulate from directly, Galichon (2022) show that it admits a convenient representation in terms of a positive stable random variable, which also gives a constructive existence result. Let z be positive stable with index λ, denoted $z \sim \mathcal{P}(\lambda)$, meaning that its Laplace transform satisfies

$$\mathbb{E}\left[e^{-tz}\right] = e^{-t^\lambda}, \qquad t \geq 0. \tag{2.5}$$

Then the transformation

$$\nu \equiv \lambda \log z \tag{2.6}$$

produces a Cardell draw, $\nu \sim C(\lambda)$. This representation is convenient computationally because it reduces Cardell simulation to simulation of a one-sided stable random variable.[4]

[4]Given (2.6), $\mathbb{E}[\nu] = (1-\lambda)\gamma$ and $\text{Var}(\nu) = (1-\lambda^2)\frac{\pi^2}{6}$, where $\gamma \simeq 0.58$ is the Euler–Mascheroni constant.

2.3 A general Cardell-based error structure

Definition 1 suggests a convenient variance-components representation of EV1 shocks. One can introduce dependence through shared Cardell components while preserving an extreme-value marginal by leaving a residual i.i.d. EV1 term with a common scale. We use this idea to parameterize the joint distribution of ϵ_{ijt} in panels by collecting all dependence-inducing components in a structured latent term η_{ijt} and retaining a remaining i.i.d. extreme-value component that delivers logit choice probabilities conditional on η_{ijt}.

Specifically, write the total unobserved utility shock as

$$\epsilon_{ijt} \;=\; \eta_{ijt} \;+\; \Lambda\, \varepsilon_{ijt}, \tag{2.7}$$

where $\varepsilon_{ijt} \overset{iid}{\sim}$ EV1 across (i, j, t) and $\Lambda \in (0, 1]$ is the common residual extreme-value scale passed to the i.i.d. component. The term η_{ijt} aggregates the structured (Cardell-based) components that generate correlation across alternatives and time.[5] We build η_{ijt} so that each ϵ_{ijt} remains marginally EV1, with all dependence governed by which latent components are shared and how they evolve over time.[6]

A flexible specification for the structured component is

$$\eta_{ijt} = \underbrace{\left(\zeta^P_{ij} + \tau^P_{i,g(j)}\right)}_{\eta^P_{ij}} + \underbrace{\left(\zeta^D_{ijt} + \tau^D_{i,g(j),t}\right)}_{\eta^D_{ijt}} + \underbrace{\tau^0_{i,g(j),t}}_{\eta^0_{ijt}}, \tag{2.8}$$

where: (i) η^P_{ij} is *persistent* (time-invariant) heterogeneity, (ii) η^D_{ijt} is *time-varying and serially correlated* heterogeneity, and (iii) η^0_{ijt} is a collection of *within-period nest shocks* that generate contemporaneous correlation across alternatives in the same nest, in the spirit of nested logit. The notation $g(j)$ denotes a (possibly application-specific) mapping from product j to a nest label. In practice we allow multiple, potentially overlapping groupings (e.g., brand and category); layering provides the organization that makes such overlap compatible with a single residual scale Λ (see below and Appendix A).

The terms ζ^P_{ij} and ζ^D_{ijt} are consumer–product components. The persistent part ζ^P_{ij} captures

[5] This structure also allows for correlation across individuals. For the remainder of the paper we focus on correlation across products and time.

[6] In the Cardell construction, the degenerate case corresponds to $\lambda = 1$, for which the associated Cardell component collapses to zero. Hence, setting all Cardell (layer) parameters to one implies $\eta_{ijt} \equiv 0$ and $\Lambda = 1$, recovering the classic multinomial logit (McFadden, 1973). Setting only the *within-period nest* parameters below one (with all persistent and dynamic parameters equal to one) yields the familiar nested logit model (McFadden, 1978).

stable match quality (e.g., long-run household–product affinity). The dynamic part ζ^D_{ijt} captures serially correlated taste shocks or latent state variables that affect utility over multiple occasions (e.g., drifting tastes). The terms $\tau^P_{i,g(j)}$, $\tau^D_{i,g(j),t}$, and $\tau^0_{i,g(j),t}$ are consumer–nest components shared by all products in nest $g(j)$. They generate cross-alternative dependence because products in the same nest share the same latent shock. The key distinction is temporal: $\tau^P_{i,g(j)}$ is time-invariant, $\tau^D_{i,g(j),t}$ is time-varying with serial dependence, and $\tau^0_{i,g(j),t}$ is purely within-period (it affects the substitution pattern at time t but does not itself propagate over time).

Our construction relies on the Cardell conjugacy property. By recursively adding scaled Cardell components and passing forward the remaining EV1 scale, one can build rich variance-components structures for η_{ijt} while keeping the marginal distribution of ϵ_{ijt} in (2.7) normalized to EV1.

Layering intuition Layering is an approach for organizing many variance components, including *disjoint* and *overlapping* nests, without losing the single-scale representation in (2.7). Formally, we introduce an ordered collection of K *layers*, indexed by $\ell = 1, \ldots, K$, where each layer ℓ partitions the alternatives into disjoint nests and assigns each product j to exactly one nest $g_\ell(j)$ in that layer. A product can load on multiple groupings by appearing in multiple layers (e.g., one layer for brands, another for categories). This ensures that every (i, j, t) receives exactly one Cardell contribution per layer, so the remaining EV1 scale passed to the i.i.d. term stays common across alternatives.

Within a layer, different nests may have different Cardell parameters (and hence different "amounts" of shared variation). To preserve a common residual scale for all j, we define a layer minimum parameter, $\underline{\lambda}_\ell$, and add an adjustment shock so that the *total* layer contribution has the same Cardell scale for every alternative, even though the *shared* part differs by nest. For dynamic layers we also require a common maximum lag length within the layer; this guarantees that dynamic variance components consume the same amount of EV1 scale for every alternative. Appendix A provides additional details and shows how this construction delivers a single residual scale Λ in (2.7).

This layered representation is also a parsimonious way to target correlation patterns. Rather than parameterizing a large covariance matrix and maintaining positive definiteness as additional sources of dependence are introduced, the researcher specifies economically interpretable shared components at the product, nest, and time level. In that sense, the Cardell construction provides a modular variance-components approach for dependence that is especially useful when persistent, transitory, and overlapping group effects must be combined in a single specification.

Layering representation. A convenient representation writes η_{ijt} as a sum of layerwise contributions,

$$\eta_{ijt} = \sum_{\ell=1}^{K} \Lambda_{\ell-1} \sum_{s=0}^{L_\ell} \underline{\lambda}_\ell^s \, \widetilde{\nu}^{(\ell)}_{i,g_\ell(j),t-s}, \tag{2.9}$$

where $L_\ell \geq 0$ is the layer-specific maximum lag (with $L_\ell = 0$ for purely persistent or purely within-period layers), $\underline{\lambda}_\ell$ is the minimum Cardell parameter in layer ℓ, $\Lambda_{\ell-1}$ is the residual EV1 scale passed forward from earlier layers, and $\widetilde{\nu}^{(\ell)}$ is a layer-ℓ adjusted Cardell primitive (all defined in Appendix A). Equation (2.9) is an accounting identity. The different subsets of layers feed ζ^P, τ^P, ζ^D, τ^D, and τ^0 in (2.8), and can be constructed sub-setting the layers in (2.8) (see Appendix A for the specific constructions of ζ^P, τ^P, ζ^D, τ^D, and τ^0).

Outside option. Let $j = 0$ denote the outside option and normalize its mean utility to $\delta_{0t} \equiv 0$. To preserve the same conditional scale Λ for all alternatives, we represent the outside-option shock as

$$u_{i0t} = \epsilon_{i0t} = \nu_{i0t} + \Lambda \, \varepsilon_{i0t}, \qquad \nu_{i0t} \overset{iid}{\sim} C(\Lambda), \tag{2.10}$$

independent of the inside-good components unless otherwise specified. In our applications we assume $\{\nu_{i0t}\}$ is i.i.d. over t, but one could analogously introduce persistence or shared components for the outside option if needed.

2.4 Choice probabilities and simulation-based likelihood

Since the representation (2.7) preserves logit choice probabilities after conditioning on the latent component, conditional on η_{ijt}, the only remaining randomness is the scaled i.i.d. EV1 term $\Lambda\varepsilon_{ijt}$, so conditional choice probabilities take the multinomial logit form with scale Λ:

$$\Pr(j_{it} = j \mid \{\eta_{ikt}, \delta_{kt}\}_{k \in \mathcal{J}}) = \frac{\exp\left(\frac{\delta_{jt}+\eta_{ijt}}{\Lambda}\right)}{\sum_{k \in \mathcal{J}} a_{kt} \exp\left(\frac{\delta_{kt}+\eta_{ikt}}{\Lambda}\right)} \tag{2.11}$$

where $j_{it} \in \{0, 1, \ldots, J\}$ denotes consumer i's realized choice at time t.

This conditional logit structure is the key computational advantage of the Cardell-based specification: correlation of the latent utility components across products and time enters only through η_{ijt}. The scale parameter Λ enters exactly as in standard logit: smaller Λ increases the effective dispersion of $\delta_{jt} + \eta_{ijt}$ and amplifies substitution among alternatives that share common

latent components through η_{ijt}.

Throughout the paper, we evaluate likelihood and related objects by simulating only the component η_{ijt} (constructed from the Cardell primitives in (2.9)) and integrating out the remaining i.i.d. type-I extreme value term $\Lambda\varepsilon_{ijt}$ in closed form.

Advantages of layered Cardell construction This construction also brings several advantages relative to Gaussian utility-error specifications, such as panel multinomial probit with serially correlated errors. First, as stated above, once we condition on the structured component η_{ijt}, we retain exact logit choice probabilities conditional on the simulated components. Second, the Cardell normalization fixes the marginal scale of the *total* shock ϵ_{ijt}: by construction each $\epsilon_{ijt} = \eta_{ijt} + \Lambda\varepsilon_{ijt}$ remains marginally EV1. This normalization keeps the parameters in δ_{jt} on a stable utility scale as we enrich the dependence structure in η_{ijt}, and it avoids scale drift that can otherwise weaken numerical conditioning when one changes the covariance specification. Third, the Cardell recursion delivers a useful *closure* (or "continuation") property. For any cutoff layer $\bar{\ell}$, the tail contribution

$$R_{ijt}^{(\bar{\ell})} \equiv \sum_{\ell=\bar{\ell}}^{\mathcal{L}} \eta_{ijt}^{(\ell)} + \Lambda\,\varepsilon_{ijt} \tag{2.12}$$

has an EV1 marginal with the continuation scale implied by the recursion after layers $1, \ldots, \bar{\ell}-1$. Shared components in the tail can still induce dependence across alternatives and time, so $R_{ijt}^{(\bar{\ell})}$ need not be i.i.d. across j, but the extreme-value closure still lets us integrate out blocks of lower-level primitives analytically once we condition on a reduced set of simulated components. We exploit this idea to construct likelihood objects with analytic choice probabilities after partial integration (for example, in our pairwise logit likelihood), and we use the same approach to compute post-estimation objects such as market shares by averaging closed-form per-occasion choice probabilities over simulated shared components. Layering plays an important practical role here: it keeps the relevant continuation scales common across alternatives, which makes these partial-integration steps well defined. Gaussian constructions do not generally offer an analogous closure: when we integrate out subsets of normal components, we typically lose the logit/nested-logit choice probabilities and must rely on higher-dimensional simulation or additional numerical integration.

The same computational advantage carries over to post-estimation objects. Market shares, elasticities, posterior choice forecasts, and product-removal counterfactuals can be computed by averaging closed-form per-occasion logit probabilities (or inclusive-value expressions) over the

simulated shared components, rather than recomputing high-dimensional Gaussian probabilities inside each counterfactual loop. This is particularly valuable in applications, like ours, that require repeated probability evaluations for prediction and welfare analysis.

Simulated Joint Likelihood. In principle, the model can be estimated by simulation over the Cardell components. Fix the number of simulation draws S. For each draw $s = 1, \ldots, S$, one could simulate the primitive Cardell shocks that enter (2.9) and construct $\{\eta_{ijt}^{[s]}\}_{i,j,t}$. Conditional on $\eta^{[s]}$, one can compute (2.11) exactly. The simulated likelihood contribution for consumer i is then

$$\ell_i(\theta) \;=\; \frac{1}{S}\sum_{s=1}^{S}\prod_{t=1}^{T}\Pr\left(j_{it} \mid \eta_{it}^{[s]}\right), \tag{2.13}$$

where θ collects the parameters in δ_{jt}, the parameters governing the Cardell components, and any parameters determining Λ. The sample log-likelihood is $\mathcal{L}(\theta) = \sum_{i=1}^{N} \log \ell_i(\theta)$. Relative to a standard approach with Gaussian errors, the simulation burden is reduced because we do not simulate the idiosyncratic shock at the alternative–time level.

In long panels, however, the simulated likelihood in (2.13) can be numerically fragile and sensitive to Monte Carlo error when S is fixed. Section 4.3 develops a pairwise composite-likelihood approach that mitigates these issues while taking advantage of the fact that (2.11) can be computed exactly conditional on $\eta^{[s]}$.

2.5 Simulation of Cardell shocks

Implementing (2.13) requires draws from $C(\lambda)$ at many candidate values of λ. Because λ changes the shape of the Cardell distribution, draws cannot be reused across parameter values by a simple rescaling. While Cardell (1997) characterizes the family through an EV1-preserving convolution property and uses it to derive analytic probability expressions, that characterization does not by itself yield a convenient simulation algorithm. We rely on the representation in Galichon (2022) linking $C(\lambda)$ to the log-stable distribution, which permits simulation by transforming draws from simple base distributions (uniform and exponential) using standard stable-sampling algorithms.

Let $v \sim \text{Unif}(0, \pi)$ and $e \sim \text{Exp}(1)$ be independent, and define

$$z \;=\; \frac{\sin(\lambda v)}{(\sin v)^{1/\lambda}}\left(\frac{\sin((1-\lambda)v)}{e}\right)^{\frac{1-\lambda}{\lambda}}. \tag{2.14}$$

Then $\nu = \lambda \log z \sim C(\lambda)$.[7] Holding fixed the underlying (v, e) draws across candidate values of λ yields common random numbers, which reduce Monte Carlo noise and smooth the simulated objective.

When repeated evaluation of (2.14) is costly, we precompute an approximation to the inverse CDF $Q(\lambda, p) = F_\lambda^{-1}(p)$ on a grid of λ values and interpolate in (λ, p). Estimation then draws $p \sim \text{Unif}(0, 1)$ and sets $\nu \approx \widehat{Q}(\lambda, p)$. This replaces repeated stable transforms with inexpensive lookup and interpolation.

The remainder of the paper develops a set of economically interpretable special cases of (2.7)–(2.9) and discusses identification and estimation in long micro panels.

3 Specializing the Framework

This section specializes the general Cardell decomposition by adapting the structure of $\eta_{ijt} = \{\eta_{ij}^P, \eta_{ijt}^D, \eta_{ijt}^0\}$ in (2.7) and the scale Λ of the remaining i.i.d. extreme-value component. We show that this framework encompasses both standard discrete choice models (e.g. nested logit) with known analytic likelihoods, and discrete choice models with novel error structures without known analytic likelihoods.

3.1 Static within-period correlation

3.1.1 Single layer nested logit

A natural baseline is correlation across products within a purchase occasion, without persistence over time. Let the inside products be partitioned into mutually exclusive nests $g \in \mathcal{G}$, and let $g(j)$ denote the nest containing product j.[8] We treat the outside option as belonging to its own singleton nest; this ensures that all alternatives can be handled within a single representation.[9]

Fix nest-specific parameters $\lambda_g \in (0, 1]$ for each $g \in \mathcal{G}$, and define $\underline{\lambda} \equiv \min_{g \in \mathcal{G}} \lambda_g$. For each consumer i, product j, and period t, draw mutually independent primitives

$$\nu_{i,g(j),t} \sim C(\lambda_{g(j)}) \qquad \text{and} \qquad \nu_{ijt}^{\text{adj}} \sim C\left(\frac{\underline{\lambda}}{\lambda_{g(j)}}\right),$$

[7] This transform is due to Kanter (1975) and Chambers et al. (1976). In numerical work, we evaluate (2.14) in log space and bound λ away from 0 and 1.

[8] As discussed in Sections 2.3 and Appendix A , we can accommodate arbitrary nests with layering, which requires indexing the nests by layer $g_l(j)$. In this example we have a single level of mutually exclusive nests, so our notation can simplify $g_1(j) = g(j)$.

[9] In practice we can represent the outside option utility as $\epsilon_{i0t} = \nu_{i0t} + \Lambda\varepsilon_{i0t}$ where $\nu_{i0t} \overset{iid}{\sim} C(\Lambda)$.

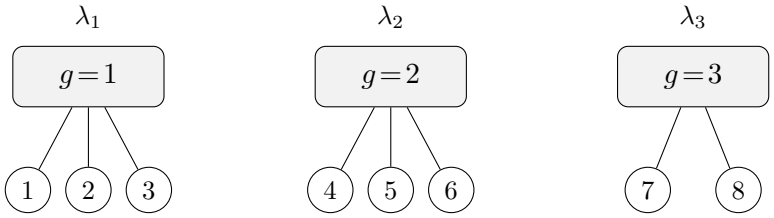


Figure 3.1: Single-level nesting. Each product belongs to exactly one nest $g(j) \in \{1, 2, 3\}$. The nest-level Cardell shock $\nu_{i,g(j),t}$ is shared by all products in the same nest, and the layering procedure ensures a common $\Lambda = \underline{\lambda}$ across all products.

where $\nu_{i,g(j),t}$ is shared across all products in nest $g(j)$ and ν^{adj}_{ijt} is product-specific. By the the composition rule in (2.4), their scaled sum satisfies $\nu_{i,g(j),t} + \lambda_{g(j)}\,\nu^{\text{adj}}_{ijt} \sim C(\underline{\lambda})$ for every product j, regardless of which nest it belongs to. This preserves a common scale $\Lambda = \underline{\lambda}$ across all products.[10]

Mapping the single-layer nested logit into our general framework requires:

$$\eta^P_{ij} = 0, \qquad \eta^D_{ijt} = 0, \qquad \eta^0_{ijt} = \underbrace{\nu_{i,g(j),t} + \lambda_{g(j)}\,\nu^{\text{adj}}_{ijt}}_{\tau^0_{ijt}}, \qquad \Lambda = \underline{\lambda}. \tag{3.1}$$

The total stochastic term is then

$$\epsilon_{ijt} = \nu_{i,g(j),t} + \lambda_{g(j)}\,\nu^{\text{adj}}_{ijt} + \underline{\lambda}\,\varepsilon_{ijt}, \qquad \varepsilon_{ijt} \overset{iid}{\sim} \text{EV1}, \tag{3.2}$$

which is marginally EV1 by conjugacy (Definition 1). Products in the same nest share $\nu_{i,g(j),t}$ and are therefore correlated, while products in different nests are conditionally independent. In this specification, $\Lambda = \underline{\lambda}$; when nesting is later combined with persistent or dynamic components, Λ will incorporate additional parameters. This construction of ϵ_{ijt} connects directly to the standard nested logit. By conjugacy, $\nu^{\text{adj}}_{ijt} + (\underline{\lambda}/\lambda_{g(j)})\,\varepsilon_{ijt} \sim \text{EV1}$, independently across products. Conditional on the nest shock $\nu_{i,g(j),t}$, each product's remaining error is i.i.d. EV1, which is exactly the standard nested-logit structure. Evaluating the simulated likelihood from (3.2) via the conditional logit kernel (2.11) recovers the standard nested-logit choice probability (Cardell, 1997), within simulation error. To make the connection to our notation explicit, define the within-nest inclusive value

$$IV_{gt} \;\equiv\; \log\left(\sum_{k:\,g(k)=g,\;a_{kt}=1} \exp\!\left(\frac{\delta_{kt}}{\lambda_g}\right)\right). \tag{3.3}$$

[10] When $\lambda_g = \underline{\lambda}$, the adjustment draw is distributed $C(1) = 0$ everywhere, so only the nest-level shock remains. In estimation, which nest attains the minimum is determined by the parameter values.

Then the standard nested-logit choice probability, parameterized by the same λ_g terms, is

$$\Pr(j_{it} = j) \;=\; \underbrace{\frac{\exp\big(\delta_{jt}/\lambda_{g(j)}\big)}{\sum_{k:\, g(k)=g(j),\, a_{kt}=1} \exp\big(\delta_{kt}/\lambda_{g(j)}\big)}}_{\text{within-nest}} \cdot \underbrace{\frac{\exp\big(\lambda_{g(j)}\, IV_{g(j),t}\big)}{\sum_{h\in\mathcal{G}} \exp\big(\lambda_h\, IV_{ht}\big)}}_{\text{across nests}}, \tag{3.4}$$

where the sum in the denominator runs over all nests, including the outside option's singleton nest. In the single-layer, purely static case, this closed form is typically preferable to simulation. The simulated Cardell representation (3.2) becomes useful when one combines within-period nesting with persistent match effects or serial dependence in unobservables; in those richer specifications, the analytic expression (3.4) is no longer available, but the conditional logit structure (2.11) remains, and estimation proceeds by simulating only the Cardell components.[11]

3.1.2 Hierarchical tree-type nesting

The single-layer construction and its connection to the analogous closed form GEV version of nested logit extends to hierarchical nesting with multiple levels, provided the nests at each level are mutually exclusive. In the layering framework of Appendix A, each nesting level corresponds to a layer l with lag length $L_l = 0$ (within-period only), its own partition of products into disjoint nests $\mathcal{G}_l$, and nest-specific parameters $\lambda_{lg} \in (0, 1]$.

As a hierarchical example, consider two levels: an outer grouping $g_1(j)$ (e.g., product category) and an inner grouping $g_2(j)$ (e.g., brand within category), where each inner nest is contained entirely within one outer nest. Each layer contributes an adjusted Cardell shock $\widetilde{\nu}^{(l)}_{i,g_l(j),t} \sim C(\underline{\lambda}_l)$ to every product in nest $g_l(j)$, where $\underline{\lambda}_l \equiv \min_{g\in\mathcal{G}_l} \lambda_{lg}$ is the layer minimum. The layering procedure (Appendix A) ensures that this contribution has a common Cardell scale across all products within each layer, regardless of which nest the product belongs to.

Mapping into the general framework:

$$\eta^P_{ij} = 0, \qquad \eta^D_{ijt} = 0, \qquad \eta^0_{ijt} = \underbrace{\widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1\, \widetilde{\nu}^{(2)}_{i,g_2(j),t}}_{\tau^0_{ijt}}, \qquad \Lambda = \underline{\lambda}_1\, \underline{\lambda}_2. \tag{3.5}$$

[11] Estimating the nested logit via (2.11) allows for a full information approach. When the analytic full-information likelihood is difficult to optimize, the simulated approach offers a full-information alternative to the two-step estimators commonly used in practice, which are consistent but asymptotically inefficient.

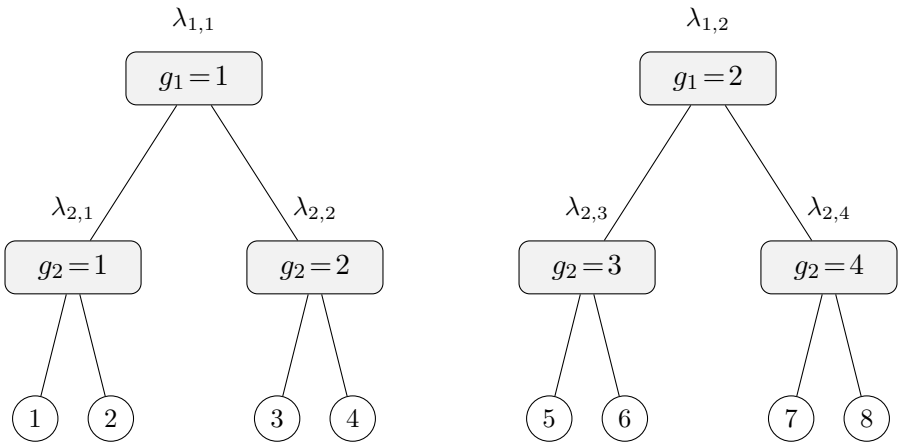


Figure 3.2: Hierarchical (multi-level) nesting. Each inner nest $g_2(j)$ is contained entirely within one outer nest $g_1(j)$. Layer 1 (outer) contributes $\widetilde{\nu}^{(1)}_{i,g_1(j),t}$ and layer 2 (inner) contributes $\widetilde{\nu}^{(2)}_{i,g_2(j),t}$. Products sharing an inner nest are more strongly correlated than products sharing only an outer nest. As with single-level nesting, this recovers the analytic multi-level nested-logit probability.

The total stochastic term is

$$\epsilon_{ijt} = \widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1 \, \widetilde{\nu}^{(2)}_{i,g_2(j),t} + \underline{\lambda}_1 \, \underline{\lambda}_2 \, \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \overset{iid}{\sim} \text{EV1}, \tag{3.6}$$

which is marginally EV1 by iterated conjugacy. The correlation pattern respects the hierarchy: products sharing an inner nest g_2 are more strongly correlated than products sharing only an outer nest g_1.

As with the single-layer case, evaluating the simulated likelihood via the conditional logit kernel (2.11) recovers the standard multi-level nested-logit choice probability within simulation error. One can construct deeper hierarchies by adding layers, with each additional level contributing one adjusted Cardell draw per nest and one multiplicative factor $\underline{\lambda}_l$ to Λ.

Additionally, the layering framework allows a hybrid strategy: one can simulate layers sequentially until the remaining tail has a disjoint nesting structure, at which point the rest can be evaluated analytically via the continuation property (Appendix A.1). For example, in a hotel setting one might simulate draws for a beachfront layer, and conditional on those draws, evaluate the remaining quality-tier hierarchy analytically. The simulation dimension is determined by how many layers need to be resolved before the remainder forms a tree, not by the total nest structure.

3.2 Overlapping Nests

The specifications in Sections 3.1–3.1.2 require mutually exclusive nests at each level, which matches the standard nested logit and multi-level nested logit models. Many applications, however, call for overlapping group structures in which a product belongs to multiple nests simultaneously. For example, a hotel might belong to both a "waterfront" nest and a "luxury" nest, generating correlation with other waterfront properties through one channel and with other luxury properties through another.

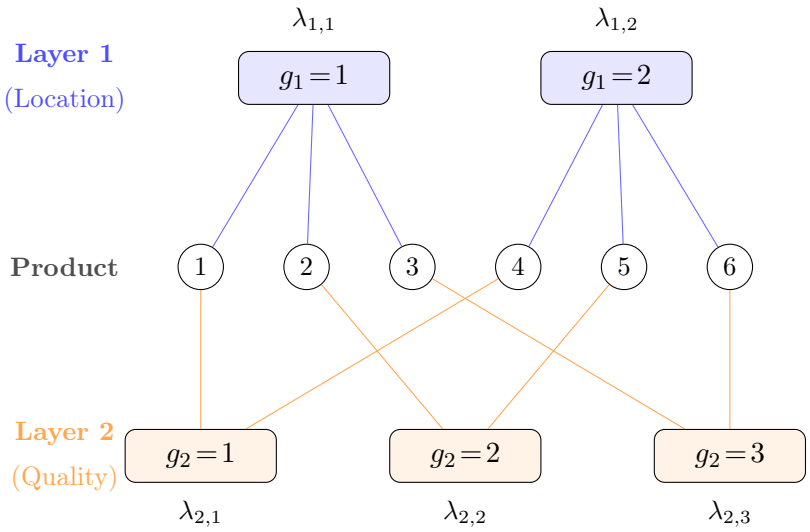


Figure 3.3: Overlapping nests via layering. Products are partitioned by location (layer 1, top) and by quality tier (layer 2, bottom). Within each layer, nests are disjoint, but across layers the partitions need not align: product 1 ($g_1 = 1$, $g_2 = 1$) shares a location shock with products 2 and 3 and a quality shock with product 4 ($g_1 = 2$, $g_2 = 1$). The layering procedure guarantees a common Λ across all products despite the overlapping structure.

Existing GEV-based approaches to overlapping nests include the paired combinatorial logit, the cross-nested logit (Vovsha, 1997), and the generalized nested logit (GNL; Wen and Koppelman, 2001). The GNL assigns each product j an allocation parameter $\alpha_{jk} \geq 0$ for each nest k, with $\sum_k \alpha_{jk} = 1$, reflecting the degree to which j belongs to nest k. While the GNL provides an analytic choice probability, the product-specific allocation weights mean that the variance contribution of a given nest generally differs across products within that nest. As a result, the split between structured (nest) and idiosyncratic components is product-specific, creating implicit heteroskedasticity in the residual scale even though the total error is marginally EV1 for each product.

The Cardell construction with layering offers an alternative that avoids this issue. In our framework, overlapping nests are accommodated by assigning each grouping dimension to a separate

layer. Within each layer l, the nests are disjoint, and the layering procedure produces adjusted Cardell shocks $\widetilde{\nu}^{(l)}_{i,g_l(j),t} \sim C(\underline{\lambda}_l)$ with a common scale for every product, regardless of nest membership. Across layers, nests may overlap: the same product can belong to different nests in different layers, generating the desired cross-cutting correlation structure.

For example, suppose products are grouped by location ($l = 1$) and by quality tier ($l = 2$), with each grouping forming its own layer of disjoint nests. A waterfront luxury hotel shares a location shock with other waterfront properties and a quality shock with other luxury properties. Mapping into the general framework:

$$\eta^P_{ij} = 0, \qquad \eta^D_{ijt} = 0, \qquad \eta^0_{ijt} = \underbrace{\widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1\, \widetilde{\nu}^{(2)}_{i,g_2(j),t}}_{\tau^0_{ijt}}, \qquad \Lambda = \underline{\lambda}_1\, \underline{\lambda}_2. \tag{3.7}$$

The total stochastic term is

$$\epsilon_{ijt} = \widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1\, \widetilde{\nu}^{(2)}_{i,g_2(j),t} + \Lambda\, \varepsilon_{ijt}, \qquad \varepsilon_{ijt} \overset{iid}{\sim} \text{EV1}, \tag{3.8}$$

which is marginally EV1 by conjugacy. Products sharing a layer-1 nest (e.g., same location) are correlated through $\widetilde{\nu}^{(1)}$; products sharing a layer-2 nest (e.g., same quality tier) are correlated through $\widetilde{\nu}^{(2)}$; products sharing nests in both layers have stronger correlation.

Importantly, the residual scale $\Lambda = \underline{\lambda}_1\, \underline{\lambda}_2$ is common across all products by construction. This is a structural difference from the GNL and related models, where the product-specific allocation parameters α_{jk} mean that different products can have different effective residual scales. In the Cardell approach, the researcher defines *which* products share correlation (the nest structure across layers) and conditional on the chosen layer structure, the λ_{lg} govern the strength of within-period correlation. The layering procedure handles the variance accounting automatically.[12]

Because the Cardell construction enforces uniform within-nest contributions and a common Λ, it does not recover the GNL choice probability when nests overlap. The two approaches impose different variance structures and generate different substitution patterns, even when targeting the same qualitative pattern of overlapping membership. However, the conditional logit kernel (2.11) applies uniformly across all products in the Cardell construction, and estimation proceeds by simulating the nest-level and adjustment shocks.

[12] Section 2.3 and Appendix A provide the general construction for arbitrary overlapping structures, including the pairwise case. Additional layers are accommodated by extending the scale recursion $\Lambda = \prod_l \underline{\lambda}_l$.

3.3 Persistent match effects in panels

The preceding specifications generate correlation across products within a choice occasion, this section introduces correlation across time. In many settings, such as household panels, researchers observe households making repeated choices over time, and those choices may reflect persistent latent affinity for particular brands or varieties even after controlling for observed characteristics and prices. Unlike the familiar within-period nesting structures, this form of temporal dependence has no comparable closed-form GEV analogue. However, Cardell construction handles this kind of across time dependence by introducing a time-invariant component at the (i, j) level. Fix $\lambda_P \in (0, 1]$, draw

$$\nu_{ij} \sim C(\lambda_P) \tag{3.9}$$

independently across (i, j), and set

$$\eta^P_{ij} = \nu_{ij}, \qquad \eta^D_{ijt} = 0, \qquad \eta^0_{ijt} = 0, \qquad \Lambda = \lambda_P. \tag{3.10}$$

Then

$$\epsilon_{ijt} = \nu_{ij} + \lambda_P \varepsilon_{ijt}, \tag{3.11}$$

which is marginally EV1 by conjugacy (Definition 1). Because the same ν_{ij} enters every period, the dependence between any two trips for a given (i, j) pair is constant regardless of the time gap between them. This contrasts with the AR(1) specification in Section 3.4, where dependence decays with the distance between trips.

Conditional on the match effects $\{\nu_{ij}\}_j$, single-period choice probabilities retain the standard logit form via (2.11). The persistent draws induce dependence across periods for a given consumer–product pair, which enters the joint and pairwise likelihoods developed in Section 4.3.

Persistent Cardell match effects serve a similar role to random coefficients on product fixed effects, capturing unobserved consumer–product affinity that is stable over time. The difference is that persistent match effects are embedded in the Cardell error structure rather than in the coefficient vector, which preserves the conditional logit kernel and offers estimation advantages discussed in Section 4.3.

Extensions. The persistent specification admits several natural generalizations within the same framework. First, one can add persistence at the nest level rather than (or in addition to) the product

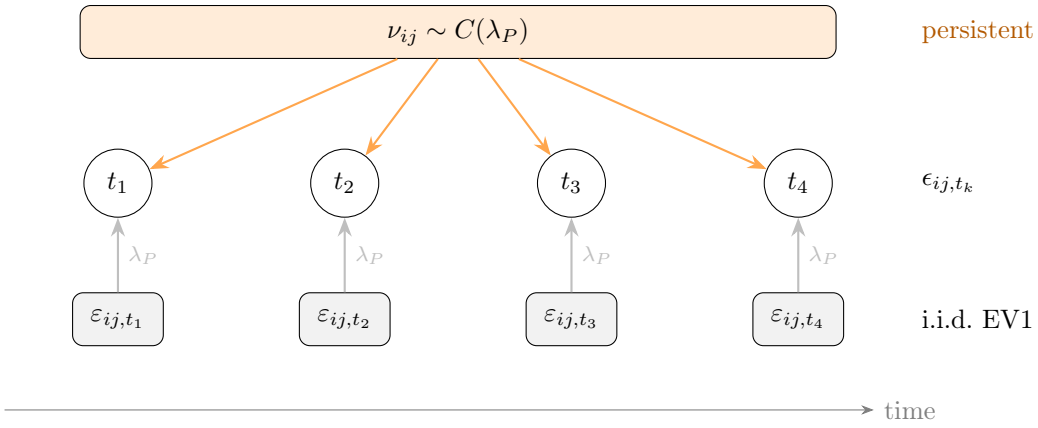


Figure 3.4: Persistent match effects for a given consumer–product pair (i, j). A single time-invariant Cardell draw $\nu_{ij} \sim C(\lambda_P)$ enters every trip with equal weight, generating stable serial correlation. Each trip also receives an i.i.d. EV1 shock scaled by λ_P. The total $\epsilon_{ij,t_k} = \nu_{ij} + \lambda_P \varepsilon_{ij,t_k}$ is marginally EV1 at every trip. Unlike the within-period nesting figures, which show the correlation structure across products, this figure depicts the time-series structure for a single product.

level, so that a consumer's affinity for a product category or brand persists across trips. Second, the parameter λ_P can be made nest- or product-specific, allowing different groups of products to have different persistent variance shares. Third, for applications that call for it, λ_P can interact with consumer demographics or be treated as a random coefficient, generating heterogeneity in the degree of persistence across consumers. In each case, the layering procedure (Appendix A) ensures that the resulting error structure preserves EV1 marginals and a common Λ across products.

3.4 AR(1) Cardell shocks

While (3.11) captures dependence through stable heterogeneity, other applications instead call for serial correlation in transitory taste shocks. In such settings, a consumer's utility from a product at one date is correlated with utility at nearby dates, with the correlation decaying over time. For example, a household may face a temporary but persistent consumption need that raises the utility of a subset of products for several periods before dissipating.

We construct this dependence by writing the dynamic component of the error as a weighted sum of independent Cardell primitives. The conjugacy property (Definition 1) ensures that the full match quality term remains EV1. We then apply the composition rule in (2.4) to collapse sums of primitives into single Cardell-distributed blocks, yielding an observed-trip representation that separates the component of the shock newly realized at any date from the component inherited from earlier dates.

3.4.1 Recursive specification

Let t index calendar time and let t_0 denote a pre-sample initialization date.[13] Define the dynamic structured component by the recursion

$$\eta^P_{ij} = 0, \qquad \eta^D_{ijt} = \underbrace{\nu_{ijt} + \overbrace{\sum_{t'=t_0+1}^{t-1} \lambda_D^{t-t'} \nu_{ijt'}}^{\lambda_D \eta^D_{ij,t-1}}}_{\zeta^D_{ijt}}, \qquad \eta^0_{ijt} = 0, \qquad \Lambda = \lambda_D^{t-t_0}, \tag{3.12}$$

with $\nu_{ijt} \overset{iid}{\sim} C(\lambda_D)$ and $\lambda_D \in [0, 1)$. In the fully recursive specification, t_0 is taken far into the past so that $\Lambda = \lambda_D^{t-t_0} \approx 0$ for all observed dates; the dynamic structure carries essentially all of the variance and the residual EV1 term is negligible. In the finite-memory variant (see below), $t_0 = t - L$ for finite L, so that $\Lambda = \lambda_D^L > 0$ and the residual EV1 term remains active.

Calendar-time representation. Iterating (3.12) forward from t_0 yields

$$\epsilon_{ijt} \;=\; \underbrace{\nu_{ijt} \;+\; \overbrace{\sum_{t'=t_0+1}^{t-1} \lambda_D^{t-t'}\, \nu_{ijt'}}^{\lambda_D\, \zeta^D_{ij,(t_0,\,t-1]}}}_{\eta^D_{ijt} \,\equiv\, \zeta^D_{ij,(t_0,t]}} \;+\; \lambda_D^{t-t_0}\, \varepsilon_{ijt}, \tag{3.13}$$

with $\varepsilon_{ijt} \overset{iid}{\sim}$ EV1, independent of the Cardell primitives.[14] As with our other constructions, $\epsilon_{ijt} \sim$ EV1 via conjugacy (Definition 1). From the primitives ν and λ_D, we now introduce an accumulation term, $\zeta^D_{ij,(a,b]}$, which we define as:

$$\zeta^D_{ij,(a,b]} \;\equiv\; \sum_{t'=a+1}^{b} \lambda_D^{b-t'}\, \nu_{ijt'} \;\sim\; C\big(\lambda_D^{b-a}\big). \tag{3.14}$$

Notably, the accumulation term $\zeta^D_{ij,(a,b]}$ is itself Cardell distributed (via the composition rule in (2.4)). Because accumulations over non-overlapping intervals are built from distinct primitives, they are independent. This means we can decompose ϵ_{ijt} into independent, trip-to-trip Cardell blocks rather than tracking every calendar period as in (3.13).

[13] In the fully recursive version, one can think of t_0 as $t - \infty$.

[14] Again, in the fully recursive version, $\lambda_D^{t-t_0} \to 0$, so $\lambda_D^{t-t_0}\, \varepsilon_{ijt} \to 0$ and is ignorable.

Observed-trip representation. Consumers are often observed at irregular intervals, making the calendar-time representation impractical for estimation, as it requires simulated draws for every consumer period, meaning the dimension of simulation grows with the span of the calendar panel. As an alternative, we can iteratively apply our accumulation to construct observed-trip representations of ϵ_{ijt}. Let $T_i = \{t_1 < t_2 < \cdots < t_n\}$ denote the ordered trip times for consumer i, with $t_0 < t_1$.

To illustrate the observed-trip representation, consider a consumer with three trips $T_i = \{t_1, t_2, t_3\}$. We can construct consumer $i's$ total match quality term for product j for each trip as:

$$\epsilon_{ij,t_1} = \zeta^D_{ij,(t_0,t_1]} + \lambda_D^{t_1-t_0}\,\varepsilon_{ij,t_1}, \tag{3.15}$$

$$\epsilon_{ij,t_2} = \zeta^D_{ij,(t_1,t_2]} + \lambda_D^{t_2-t_1}\,\zeta^D_{ij,(t_0,t_1]} + \lambda_D^{t_2-t_0}\,\varepsilon_{ij,t_2}, \tag{3.16}$$

$$\epsilon_{ij,t_3} = \zeta^D_{ij,(t_2,t_3]} + \lambda_D^{t_3-t_2}\,\zeta^D_{ij,(t_1,t_2]} + \lambda_D^{t_3-t_1}\,\zeta^D_{ij,(t_0,t_1]} + \lambda_D^{t_3-t_0}\,\varepsilon_{ij,t_3}. \tag{3.17}$$

In this example, each earlier interval block carries forward with increasing decay, generating serial correlation in ϵ_{ijt} across trips. More generally, for the kth observed trip at time $t_k \in T_i$, we can construct ϵ_{ij,t_k} as

$$\epsilon_{ij,t_k} \;=\; \sum_{s=1}^{k} \lambda_D^{t_k-t_s}\,\zeta^D_{ij,(t_{s-1},\,t_s]} \;+\; \lambda_D^{t_k-t_0}\,\varepsilon_{ij,t_k}, \tag{3.18}$$

where $\zeta^D_{ij,(t_{s-1},t_s]} \sim C(\lambda_D^{t_s-t_{s-1}})$, mutually independent across $s = 1, \ldots, k$.

Because $\zeta^D_{ij,(t_{k-1},t_k]}$ is itself Cardell distributed, we can simulate each $\zeta^D_{ij,(t_{k-1},t_k]}$ with a single draw from $C(\lambda_D^{t_k-t_{k-1}})$ for each simulation, so the full dynamic path at observed trip times is built from $|T_i|$ draws regardless of how many calendar periods the panel spans. It is worth emphasizing this benefit as it is common, in practice, for researchers to have long panels with infrequent consumer-trips. For example, consider a typical consumer observed making 10 trips over two years (104 weekly intervals). The observed-trip representation requires 10 dimensions of integration rather than 104.[15]

As noted above, in the fully recursive case we can drop the inter-temporal EV1 term, ε_{ijt} since $\lambda_D^{t_k-t_0} \to 0$ and $\lambda_D^{t_k-t_0}\,\varepsilon_{ij,t_k} \to 0$. At the same time, the first trip's accumulation block becomes EV1: $\zeta^D_{ij,(t_0,t_1]} \sim C(\lambda_D^{t_1-t_0}) \to C(0) = \text{EV1}$. So (3.18) reduces to

$$\epsilon_{ij,t_k} \;=\; \sum_{s=1}^{k} \lambda_D^{t_k-t_s}\,\zeta^D_{ij,(t_{s-1},\,t_s]}, \tag{3.19}$$

[15]This still understates the relative dimensions reduction since alternative calendar formats would also require additional dimensions of integration to get the initial conditions.

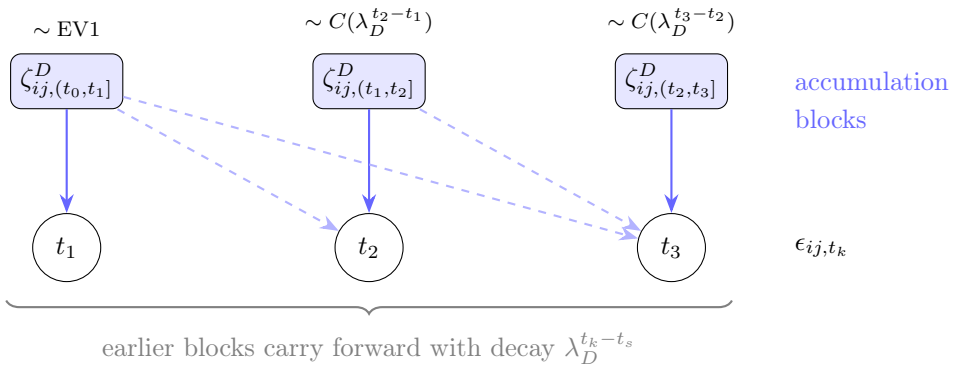


Figure 3.5: AR(1) Cardell shocks for a given consumer–product pair (i, j) (fully recursive, observed-trip representation). Each inter-trip interval produces one accumulation block $\zeta^D_{ij,(t_{k-1},t_k]}$. Solid arrows show a block entering its own trip; dashed arrows show the same block carried forward to later trips with geometric decay $\lambda_D^{t_k-t_s}$. The first block is EV1 (since $t_0 \to -\infty$); subsequent blocks are $C(\lambda_D^{t_k-t_{k-1}})$. The simulation dimension is $|T_i|$, not the calendar span.

with $\zeta^D_{ij,(t_0,t_1]} \sim$ EV1 and $\zeta^D_{ij,(t_{s-1},t_s]} \sim C(\lambda_D^{t_s-t_{s-1}})$ for $s \geq 2$.

In practice, since $\zeta^D_{ij,(t_0,t_1]} \sim$ EV1, one can avoid drawing a long sequence of terms over many pre-periods to initialize $\zeta^D_{ij,(t_0,t_1]}$, and instead draw one term from the EV1 distribution for consumer i's first trip.

3.4.2 Finite-memory special case

We can also consider a variant of the fully recursive AR(1) specification above that imposes limited memory. In this truncated variant, we replace the infinite history with a sliding window of L lags by setting the initialization date to $t_0 = t - L$ for each calendar period. The resulting specification is

$$\epsilon_{ijt} \;=\; \sum_{l=0}^{L-1} \lambda_D^l \, \nu_{ij,t-l} \;+\; \lambda_D^L \, \varepsilon_{ijt}, \qquad \Lambda = \lambda_D^L. \tag{3.20}$$

This adjustment to the specification has two notable consequences for the match term construction and implementation. First, unlike equation 3.19, we don't drop the inter-temporal EV1 term, ε_{ijt} because $\Lambda = \lambda_D^L > 0$, which preserves exact conditional logit choice probabilities at every date. Second, the memory window $[t_k - L, t_k]$ may not align with observed trip times, so the window boundary can fall in the interior of an inter-trip interval.[16]

[16] See Appendix B.1 for the trip representation.

3.5 Summary and mapping to the general framework

Table 3.1 summarizes how the specifications in this section map into the general decomposition (2.7). The unifying feature is that the dependence structure is carried by η_{ijt} through shared Cardell components, while the i.i.d. extreme-value term $\Lambda\varepsilon_{ijt}$ preserves logit choice probabilities conditional on the latent components whenever $\Lambda > 0$.

The specifications fall into three groups. Single-level and hierarchical nested logit with mutually exclusive nests recover known analytic GEV expressions within simulation error. Overlapping nests produce a related but distinct model from the GNL, with a common Λ guaranteed by layering. Persistent match effects, AR(1) Cardell shocks, and their combination generate temporal dependence for which no analytic likelihood exists; in these cases, estimation relies on the conditional logit kernel and simulation of the Cardell components.

The specifications can be combined additively in economically natural ways. In particular, the model studied in the next section combines within-period nesting, persistent match effects, and AR(1) Cardell shocks, yielding a flexible yet tractable structure for long household panels.

Table 3.1: Benchmark specifications within the generalized Cardell framework

Specification	η^P_{ij}	η^D_{ijt}	η^0_{ijt}	Λ
Analytic equivalence (mutually exclusive nests)				
Single-level nested logit	0	0	$\nu_{i,g(j),t} + \lambda_{g(j)}\nu^{\text{adj}}_{ijt}$	$\underline{\lambda}$
Hierarchical nested logit	0	0	$\widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1\,\widetilde{\nu}^{(2)}_{i,g_2(j),t}$	$\underline{\lambda}_1\underline{\lambda}_2$
Related analytic models exist but impose different variance structure				
Overlapping nests	0	0	$\widetilde{\nu}^{(1)}_{i,g_1(j),t} + \underline{\lambda}_1\,\widetilde{\nu}^{(2)}_{i,g_2(j),t}$	$\underline{\lambda}_1\underline{\lambda}_2$
No analogous analytic likelihood				
Persistent match effects	ν_{ij}	0	0	λ_P
Finite-memory AR(1)	0	$\sum_{\ell=0}^{L-1} \lambda_D^{\ell}\,\nu_{ij,t-\ell}$	0	λ_D^L
Recursive AR(1)	0	$\sum_{s=1}^{k} \lambda_D^{t_k-t_s}\,\zeta^D_{ij,(t_{s-1},t_s]}$	0	$\to 0$
Persistent + recursive AR(1)	ν_{ij}	$\lambda_P \sum_{s=1}^{k} \lambda_D^{t_k-t_s}\,\zeta^D_{ij,(t_{s-1},t_s]}$	0	$\to 0$

Notes: Each row specifies the nonzero components of $\eta_{ijt} = \eta^P_{ij} + \eta^D_{ijt} + \eta^0_{ijt}$ in (2.7). When $\Lambda > 0$, conditional on η_{ijt} the choice probabilities are multinomial logit with scale Λ; likelihood evaluation simulates only the Cardell components and integrates out the i.i.d. EV1 term in closed form. When $\Lambda \to 0$ (fully recursive specifications), the conditional-logit structure given draws is lost, but single-period choice probabilities remain standard logit since ϵ_{ijt} is still marginally EV1; simulation is needed only for the dependence component of the composite likelihood (Section 4.3). The layering procedure (Appendix A) to ensure a common Λ across products. The overlapping-nest entry has the same mathematical structure as hierarchical nesting but with cross-cutting layer partitions; see Section 3.2.

4 Estimation and Monte Carlo evidence

The preceding section showed that the generalized decomposition in Section 2 delivers a broad class of correlation structures. We now focus on a parsimonious specification—the *AR-persistent model*—that combines (i) persistent match-specific unobservables and (ii) AR(1) Cardell shocks, and we discuss practical estimation approaches for long consumer panels.

4.1 The AR-persistent specification

The benchmark cases in Section 3.5 isolate two sources of persistence in household purchase histories: stable household–product match quality and transitory shocks that decay over time. In the empirical application below both forces are potentially important. We combine a time-invariant match component as in Section 3.3 with recursive AR(1) Cardell shocks as in Section 3.4.

Mapping into the general framework, the AR–persistent specification sets

$$\eta^P_{ij} = \nu_{ij}, \qquad \eta^D_{ijt} = \underbrace{\nu_{ijt} + \overbrace{\sum_{t'=t_0+1}^{t-1} \lambda_D^{t-t'} \nu_{ijt'}}^{\lambda_D\, \eta^D_{ij,t-1}}}_{\zeta^D_{ijt}}, \qquad \eta^0_{ijt} = 0, \qquad \Lambda = \lambda_P \lambda_D^{t-t_0}, \tag{4.1}$$

with Cardell primitives $\nu_{ij} \overset{iid}{\sim} C(\lambda_P)$ and $\nu_{ijt} \overset{iid}{\sim} C(\lambda_D)$, all mutually independent. Applying the trip-level representation from Section 3.4 and taking the recursive limit ($t_0 \to -\infty$, so $\Lambda \to 0$), the total error at the kth trip for consumer i is

$$\eta_{ij,t_k} \equiv \epsilon_{ij,t_k} \;=\; \underbrace{\nu_{ij}}_{\eta^P_{ij}} + \lambda_P \underbrace{\sum_{s=1}^{k} \lambda_D^{t_k - t_s}\, \zeta^D_{ij,(t_{s-1},t_s]}}_{\eta^D_{ij,t_k}}, \tag{4.2}$$

where $\zeta^D_{ij,(t_0,t_1]} \sim$ EV1 and $\zeta^D_{ij,(t_{s-1},t_s]} \sim C\big(\lambda_D^{t_s - t_{s-1}}\big)$ for $s \geq 2$ are mutually independent accumulation blocks. The continuation scale λ_P multiplies the dynamic component because the persistent layer is ordered first in the layering recursion (Appendix A). Between consecutive trips, η^D updates via $\eta^D_{ij,t_k} = \zeta^D_{ij,(t_{k-1},t_k]} + \lambda_D^{t_k - t_{k-1}}\, \eta^D_{ij,t_{k-1}}$, so each trip introduces one new accumulation block and the transitory component decays more after longer gaps between purchases. Figure 4.1 illustrates the structure for $|T_i| = 3$. Simulating the full dynamic path requires $|T_i|$ draws regardless of the calendar span (Section 3.4).

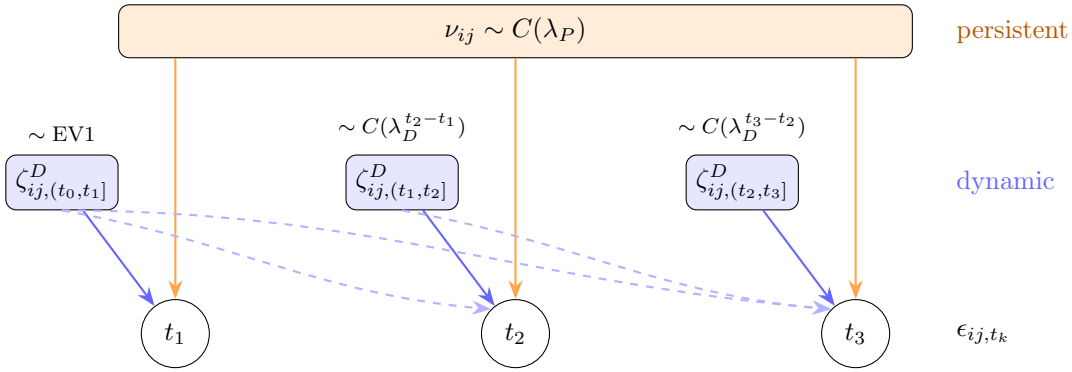


Figure 4.1: Combined persistent match effects and AR(1) Cardell shocks for a given consumer–product pair (i, j). The persistent component ν_{ij} (orange) enters every trip with equal weight. The dynamic accumulation blocks $\zeta^D_{ij,(t_{k-1},t_k]}$ (blue) each enter their own trip directly (solid) and carry forward to later trips with geometric decay $\lambda_D^{t_k - t_s}$ (dashed). Together, these generate both stable preference heterogeneity (through ν_{ij}) and decaying serial correlation (through the AR structure).

The persistent Cardell primitive ν_{ij} captures durable household–product affinity: some households consistently like particular brands or formulations more than others. The dynamic state η^D_{ij,t_k} captures transitory shocks that fade over time. Repeat purchasing can therefore arise either because a household has a persistently high match value for a product or because a recent favorable shock has not yet fully dissipated. This distinction matters for counterfactuals: the former source of persistence survives forced switching, whereas the latter naturally unwinds with time.

Truncated finite-memory benchmark. A truncated version of the AR–persistent model remains useful as a benchmark. For an integer $L \geq 0$, define

$$\epsilon^{(L)}_{ijt} = \underbrace{\nu_{ij} + \lambda_P \sum_{\ell=0}^{L-1} \lambda_D^{\ell}\, \nu_{ij,t-\ell}}_{\eta_{ijt}} + \Lambda_L\, \varepsilon_{ijt}, \qquad \Lambda_L \equiv \lambda_P\, \lambda_D^{L}, \tag{4.3}$$

with the convention that the sum is zero when $L = 0$, where the $\nu_{ij,t-\ell} \overset{iid}{\sim} C(\lambda_D)$ are Cardell primitives and $\varepsilon_{ijt} \overset{iid}{\sim}$ EV1. Conditional on the latent Cardell components, (4.3) yields exact multinomial-logit choice probabilities with scale Λ_L. We use this finite-memory specification below as a benchmark and as intuition for the recursive model, but it is not the baseline object taken to the data.

The recursive specification does not leave a positive residual EV1 term after conditioning on the latent draws. Accordingly, the draw-specific per-occasion choice probabilities are no longer exact multinomial logits. In estimation we therefore evaluate the recursive model with logit-smoothing having smoothing scale $\omega_{\text{sm}} > 0$. We treat ω_{sm} purely as a numerical device that stabilizes simulation and optimization.

4.2 Likelihood decomposition and estimation

With the error construction in hand, we can turn to estimation. We develop two related criteria: a full-sequence simulated likelihood and a pairwise likelihood designed for long panels. The Cardell construction delivers two tractability gains. First, although the total error is serially dependent across purchase occasions, it remains independent across alternatives within an occasion and is marginally EV1, so per-occasion choice probabilities are exact multinomial logit. Second, whenever a positive residual EV1 layer remains after conditioning on the structured component, draw-specific joint probabilities are again logit. The truncated AR–persistent benchmark has this property exactly;

the fully recursive specification is the limiting case with $\Lambda \to 0$, so there we use logit smoothing for the dependence term. These features let us separate the criterion into an analytic marginal term and a dependence correction handled by simulation. We first derive the decomposition and the full-sequence criterion, then show how the same structure leads to a more stable pairwise estimator for long panels.

4.2.1 Log-likelihood decomposition

Let $\theta = (\theta_\delta, \boldsymbol{\lambda})$, where θ_δ collects the parameters entering mean utility and $\boldsymbol{\lambda} = (\lambda_P, \lambda_D)$ governs the dependence structure. Fix (i, t_k). By construction, ϵ_{ij,t_k} is marginally EV1 for each j, and the underlying Cardell primitives are independent across j, so the single-period choice probabilities are standard multinomial logit:

$$\begin{aligned} p^0_{ij,t_k}(\theta_\delta) &= \frac{a_{j,t_k} \exp\big(\delta_{j,t_k}(\theta_\delta)\big)}{1 + \sum_{m=1}^{J} a_{m,t_k} \exp\big(\delta_{m,t_k}(\theta_\delta)\big)}, \quad j = 1, \ldots, J, \\ p^0_{i0,t_k}(\theta_\delta) &= \frac{1}{1 + \sum_{m=1}^{J} a_{m,t_k} \exp\big(\delta_{m,t_k}(\theta_\delta)\big)}. \end{aligned} \tag{4.4}$$

These depend only on θ_δ. Let $p^0_{i,t_k}(\theta_\delta) \equiv \sum_{j=0}^{J} y_{ij,t_k}\, p^0_{ij,t_k}(\theta_\delta)$ denote the exact logit probability of the realized outcome. The dependence parameters $\boldsymbol{\lambda}$ affect the likelihood only through intertemporal dependence in $\{j_{i,t_k}\}_{k=1}^{n_i}$.

The log-probability of consumer i's observed sequence can therefore be written as

$$\log \Pr_\theta\big(j_{i,t_1}, \ldots, j_{i,t_{n_i}}\big) \;=\; \sum_{k=1}^{n_i} \log p^0_{i,t_k}(\theta_\delta) \;+\; R_i(\theta), \tag{4.5}$$

where $R_i(\theta)$ collects the contribution of dependence across occasions. The first term is exact and depends only on θ_δ. Under the recursive AR–persistent specification, $R_i(\theta)$ is not available analytically and must be approximated by simulation.

4.2.2 Full-sequence simulated likelihood

For each simulation draw $s = 1, \ldots, S$, simulate the structured components $\eta^{[s]}_{ij,t_k}$ from the AR–persistent specification (4.2) and draw $\varepsilon^{[s]}_{i0,t_k} \sim$ EV1 for the outside option. Because the recursive specification has $\Lambda \to 0$, we evaluate the draw-specific choice kernel using logit-smoothing with scale

$\omega_{\text{sm}} > 0$:

$$
\begin{aligned}
P_{ij,t_k}^{[s]}(\theta) &= \frac{a_{j,t_k} \exp\left(\dfrac{\delta_{j,t_k}(\theta_\delta) + \eta_{ij,t_k}^{[s]}}{\omega_{\text{sm}}}\right)}{\exp\left(\dfrac{\varepsilon_{i0,t_k}^{[s]}}{\omega_{\text{sm}}}\right) + \displaystyle\sum_{m=1}^{J} a_{m,t_k} \exp\left(\dfrac{\delta_{m,t_k}(\theta_\delta) + \eta_{im,t_k}^{[s]}}{\omega_{\text{sm}}}\right)}, \quad j = 1, \ldots, J, \\
P_{i0,t_k}^{[s]}(\theta) &= \frac{\exp\left(\dfrac{\varepsilon_{i0,t_k}^{[s]}}{\omega_{\text{sm}}}\right)}{\exp\left(\dfrac{\varepsilon_{i0,t_k}^{[s]}}{\omega_{\text{sm}}}\right) + \displaystyle\sum_{m=1}^{J} a_{m,t_k} \exp\left(\dfrac{\delta_{m,t_k}(\theta_\delta) + \eta_{im,t_k}^{[s]}}{\omega_{\text{sm}}}\right)}.
\end{aligned}
\tag{4.6}
$$

Let $P_{i,t_k}^{[s]}(\theta) \equiv \sum_{j=0}^{J} y_{ij,t_k} P_{ij,t_k}^{[s]}(\theta)$ denote the probability assigned to the realized outcome at occasion t_k. The simulated sequence likelihood is

$$
\ell_i^{\text{FS}}(\theta) = \frac{1}{S} \sum_{s=1}^{S} \prod_{k=1}^{n_i} P_{i,t_k}^{[s]}(\theta), \qquad \mathcal{L}^{\text{FS}}(\theta) = \sum_{i=1}^{N} \log \ell_i^{\text{FS}}(\theta). \tag{4.7}
$$

If one instead estimates the truncated benchmark (4.3), the remaining EV1 term is positive and the draw-specific probabilities are exact conditional logits with scale Λ_L. The recursive and truncated models therefore differ only in how the draw-specific kernels are evaluated, not in the underlying economic interpretation of (λ_P, λ_D).

Computational considerations. It is useful to contrast the full sequence likelihood (4.7) and its decomposition (4.5) with a Gaussian specification that allows the same persistent and autoregressive match-quality terms. In that case, the likelihood contribution of an observed sequence $\{j_{it}\}_{t=1}^{T}$ is a multivariate normal probability over the region defined by the inequalities $u_{i\,j_{it}\,t} \geq u_{ikt}$ for all $k \neq j_{it}$ and $t = 1, \ldots, T$. Before differencing, the latent Gaussian representation has dimension on the order of $J \times (T+1)$; after the usual differencing normalization, the problem still involves a $(J-1) \times (T+1)$ integral, with a covariance matrix that is dense across t because of the persistent and autoregressive components. Likelihood evaluation therefore relies on high-dimensional simulation methods such as GHK and repeated factorizations of large covariance matrices as parameters change, making the objective increasingly noisy and numerically fragile as either J or T grows.

The Cardell specification changes the nature of the problem. Conditional on the simulated latent components that generate persistent match quality and serial dependence, the remaining idiosyncratic

shocks are i.i.d. EV1, so the conditional choice probabilities in (4.6) are exact multinomial logit. The panel likelihood is therefore an average of products of logit choice probabilities rather than a single high-dimensional multivariate normal probability. The key gain is not merely lower-dimensional simulation, but that a large share of the likelihood is handled analytically. Only the shared latent components require simulation. This substantially reduces Monte Carlo noise and yields a better-behaved criterion. In the pairwise criterion developed below, the same logic implies that the single-period term is exact logit and each dependence contribution simulates only the components shared across two occasions, which is the main computational shortcut relative to Gaussian pairwise alternatives.

4.3 Pairwise likelihood for long panels

For long panels, the full sequence likelihood can be numerically fragile because the effective dimension of the latent state grows with T and the draw-average likelihood can become dominated by a small subset of simulations.[17] We therefore implement the computational shortcut through a pairwise likelihood, a composite likelihood method that approximates joint distributions by multiplying pairs of marginal likelihoods (Varin et al., 2011). The idea is to separate the component already pinned down by exact single-period logit probabilities from the component that identifies persistence. Rather than evaluate the full sequence likelihood, we use pairwise dependence terms that simulate only the latent components shared across two purchase occasions, yielding a more stable criterion in long panels.

Our combined specification implies that the *marginal* choice probability at a given purchase occasion has a simple closed form. Fix (i,t). By construction, ϵ_{ijt} is marginally EV1 for each j and the vector $\{\epsilon_{ijt}\}_{j=0}^{J}$ is independent across alternatives because the primitive shocks $\{\nu_{ij}, \nu_{ijt}, \varepsilon_{ijt}\}$ are independent across j. Consequently, the probability that consumer i chooses product j at occasion t is the standard multinomial logit probability,

$$p_{ijt}^{0}(\theta_\delta) \;\equiv\; \Pr(j_{it} = j \mid \delta_t) \;=\; \frac{a_{jt}\exp(\delta_{jt}(\theta_\delta))}{\sum_{k:\,a_{kt}=1}\exp(\delta_{kt}(\theta_\delta))}, \tag{4.8}$$

which depends only on the parameters that enter δ_{jt}. In contrast, the dependence parameters (λ_P, λ_D) affect the likelihood only through intertemporal dependence in $\{j_{it}\}_{t=1}^{T}$.

We therefore estimate (λ_P, λ_D) using a pairwise (composite) likelihood that targets serial

[17] This issue is often referred to as weight degeneracy or particle degeneracy. See, for example, Keane (1994), Fernández-Villaverde and Rubio-Ramírez (2007), and Berry and Haile (2021).

dependence while retaining the analytic logit likelihood for θ_δ. Similar approaches have been shown to be effective in complex panel discrete choice settings where standard simulated maximum likelihood fails (Bhat, 2014; Tuzcuoglu, 2023), however the pairwise approach is particularly attractive in our framework. In a Gaussian panel model, even the single-occasion term is a multinomial-probit probability and the pair term is a multivariate normal probability, so both parts of the composite likelihood require numerical integration. Here, by contrast, the singleton term in (4.8) is exact logit and only the dependence correction needs simulation.

Using the observed-choice indicator, define the exact logit probability of the observed outcome at occasion t by

$$p^0_{it}(\theta_\delta) \equiv \sum_{j=0}^{J} y_{ijt}\, p^0_{ijt}(\theta_\delta).$$

Let $P^{[s]}_{it}(\theta)$ denote the draw-specific probability of the realized choice under logit-smoothing (4.6). For a given (i,t), define

$$\widehat{p}_{it}(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S} P^{[s]}_{it}(\theta),$$

and for any pair $t < t'$, define

$$\widehat{p}_{itt'}(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S} P^{[s]}_{it}(\theta)\, P^{[s]}_{it'}(\theta),$$

where the product inside the sum follows because, conditional on a simulated draw of the full latent path, the remaining draw-specific probabilities factor across occasions. Let $\mathcal{P}_i$ denote a set of within-consumer period pairs (potentially a random subset of all $\binom{T}{2}$ pairs). Our pairwise criterion is

$$\mathcal{PL}(\theta) \;=\; \sum_{i=1}^{N}\sum_{t=1}^{T} \log p^0_{it}(\theta_\delta) \;+\; w \sum_{i=1}^{N} \sum_{(t,t')\in\mathcal{P}_i} \log\!\left(\frac{\widehat{p}_{itt'}(\theta)}{\widehat{p}_{it}(\theta)\,\widehat{p}_{it'}(\theta)}\right). \tag{4.9}$$

The first term in (4.9) remains the exact marginal logit likelihood and depends on θ_δ. The second term is a dependence correction: it is the weighted (by w) log of the ratio between the joint probability of the realized pair and the product of the corresponding marginal probabilities. When the latent components generate no intertemporal dependence (e.g., as λ_P and λ_D approach one so that the Cardell components degenerate), $\widehat{p}_{itt'}(\theta) \approx \widehat{p}_{it}(\theta)\widehat{p}_{it'}(\theta)$, the ratio inside the logarithm is close to one, and the correction term contributes little. The weighting term, w, can be selected by the researcher. For our application, we set $w = N_{cases}/(2|\mathcal{P}|)$ where N_{cases} is the number of choice situations, and $|\mathcal{P}|$ is the number of pairs sampled.

This decomposition in (4.9) makes use of the same tractability advantage emphasized above.

Because the remaining within-period shocks are marginally i.i.d. EV1, each observed outcome contributes the exact logit probability $p^0_{it}(\theta_\delta)$, so no integration over Cardell components is needed to recover the marginal likelihood. Simulation enters only through the dependence correction, where $\widehat{p}_{itt'}(\theta)$ and $\widehat{p}_{it}(\theta)$ average over the low-dimensional latent components relevant for that pair of occasions. Relative to Gaussian pairwise alternatives, which still require repeated evaluation of Gaussian probability as parameters vary, the Cardell structure leaves within-period substitution in closed form and confines simulation to the source of serial dependence.

Identification. The structure of (4.9) clarifies the sources of identifying variation. The marginal logit term depends only on θ_δ and identifies mean-utility parameters from within-occasion variation in availability, prices, and product characteristics, exactly as in standard logit. The dependence correction depends on the full parameter vector $\theta = (\theta_\delta, \boldsymbol{\lambda})$ and provides additional information about θ_δ while being the sole source of identification for $\boldsymbol{\lambda}$, which is pinned down by within-consumer serial correlation in choices. The overall strength of serial correlation informs the scale of dependence, while the way that dependence attenuates with the gap $|t_k - t_{k'}|$ is informative about λ_D. Persistent components remain relevant even at long gaps, whereas the AR component contributes primarily at short gaps.

Table 4.1: Monte carlo design

	Value	Description
Inside products J	5	Inside options (outside option $j = 0$)
Consumers N	1,000	Per replication
Purchase occasions T	10	Potential periods per consumer
Markets M	10	Market-level price/attribute draws
Trip probability	0.9	Pr(consumer shops in t)
Availability rate	0.8	$\Pr(a_{jtm} = 1)$ for $j \geq 1$
Replications R	25	Monte Carlo replications
Estimation draws S	1024	Latent draws per consumer (scrambled Sobol)
Evaluation draws S^{eval}	256	Draws used to compute elasticities/probabilities
Pairwise pairs $\lvert\mathcal{P}\rvert$	20,000	Subsampled within-consumer period pairs
Pairwise weight w	$N_{\text{cases}}/(2\lvert\mathcal{P}\rvert)$	N_{cases} is # choice occasions
Smoothing scale ω_{sm}	0.02	Logit smoothing for $P_{ijt}^{[s]}(\theta)$
Price coefficient α	1.0	In $\delta_{mjt} = \xi_j - \alpha p_{mjt} + \beta x_{mjt}$
Characteristic coefficient β	0.2	Coefficient on x_{mjt}
Product effects ξ_j	Unif$[2, 5]$	Drawn once per replication
Persistent parameter λ_P	0.3	Match effects $\nu_{ij} \sim C(\lambda_P)$
Autocorrelation parameter λ_D	0.6	Recursive AR(1) Cardell shocks in (3.12)

4.4 Monte Carlo design

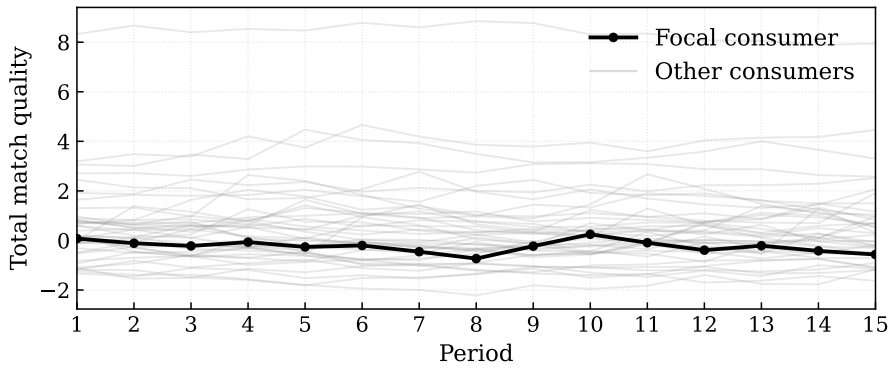
We assess the finite-sample performance of the pairwise likelihood estimator in Section 4.3 using Monte Carlo experiments. Each replication simulates a panel of $N = 1{,}000$ consumers observed over $T = 10$ potential purchase occasions, choosing among $J = 5$ inside goods and an outside option. We incorporate market-level variation by drawing $M = 10$ markets and allowing prices and observed characteristics to vary by market–product–period. In each period, consumer i has a choice occasion with probability 0.9; conditional on making a choice, each inside product $j \geq 1$ is available with probability 0.8 in market m. The outside option is always available, so the realized choice sets vary across consumers and time, similar to the scanner data we use in Section 5.

For the mean-utility component we specialize (2.2) to homogeneous tastes with a single observed characteristic,

$$\delta_{mjt} = -\alpha\, p_{mjt} + \beta\, x_{mjt} + \xi_j, \qquad \mu_{ijt} \equiv 0,$$

with a common price coefficient $\alpha > 0$, characteristic coefficient β, and product fixed effects $\{\xi_j\}$. We set $(\alpha, \beta) = (1, 0.2)$ and draw $\xi_j \overset{iid}{\sim} \text{Unif}[2, 5]$ once per replication. We draw (p_{mjt}, x_{mjt}) at the market–product–period level and introduce a modest positive correlation between prices and mean

Figure 4.2: Example of total match quality with AR-persistent error



Notes: The chart shows examples of how the total match quality term varies over 15 periods. Each line shows the match quality for one product and one consumer given the Monte Carlo parameters described in Table 4.1.

utility by shifting prices upward when ξ_j is higher.

Unobserved utility follows the *AR-persistent model*. We draw persistent Cardell primitives $\nu_{ij} \sim C(\lambda_P)$ and generate the dynamic component η^D_{ijt} via the AR(1) recursion in (3.12) with parameter λ_D and i.i.d. innovations $\nu_{ijt} \sim C(\lambda_D)$. We set $(\lambda_P, \lambda_D) = (0.3, 0.6)$ and form $\epsilon_{ijt} = \nu_{ij} + \lambda_P \eta^D_{ijt}$. To allow for outside choice, we include an i.i.d. EV1 shock in the outside option. This design preserves the extreme-value normalization marginally while inducing serial correlation within each consumer–product match.

Figure 4.2 provides an illustration of the simulated *total match-quality term* for product 1. The gray lines plot this term for a set of consumers, while the thick black line highlights a focal consumer. Two features of the Monte Carlo design are evident: (i) substantial cross-sectional heterogeneity, with some consumers exhibiting persistently high (or low) match quality for the same product, and (ii) time-series persistence within consumer, as individual trajectories evolve smoothly rather than jumping sharply from period to period.

Pairwise estimation. In each replication we estimate $(\theta_\delta, \lambda_P, \lambda_D)$ by maximizing the pairwise criterion in (4.9). The dependence term is evaluated on a random subsample of $|\mathcal{P}| = 20{,}000$ within-consumer period pairs. Let N_{cases} denote the number of observed choice occasions; we set the weight $w = N_{\text{cases}}/(2|\mathcal{P}|)$ so that the dependence correction is of comparable magnitude to the marginal logit likelihood. We compute $\widehat{p}_{it}(\theta)$ and $\widehat{p}_{itt'}(\theta)$ using $S = 1024$ common-random-number Sobol draws for the latent Cardell components. Because the recursive specification in (3.12) corresponds to the limiting case in which there are no analytic logit choice probabilities conditional on draws,

Table 4.2: Finite-sample performance: AR-persistent model estimated via pairwise likelihood

	True	Mean	Bias	Std. dev.	RMSE	95% cov.
Panel A: Structural parameters						
Price coefficient α	1.000	0.979	-0.021	0.016	0.026	1.00
Persistent Cardell λ_P	0.300	0.337	0.037	0.034	0.050	1.00
Autocorrelation Cardell λ_D	0.600	0.520	-0.080	0.180	0.193	1.00
Panel B: Elasticities and choice probabilities						
Mean own-price elasticity	-4.036	-3.938	0.098	0.152	0.120	1.00
Mean cross-price elasticity	0.687	0.674	-0.013	0.098	0.017	1.00
Mean choice probability (product 1)	0.115	0.115	0.000	0.024	0.003	1.00
Panel C: Repeat and transition probabilities						
Repeat choice probability	0.545	0.543	-0.002	0.038	0.007	1.00
Transition prob. $P_{1\to1}$	0.539	0.537	-0.003	0.046	0.009	1.00
Transition prob. $P_{1\to2}$	0.110	0.111	0.001	0.026	0.003	1.00
Transition prob. $P_{1\to3}$	0.110	0.110	0.000	0.029	0.004	1.00
Transition prob. $P_{1\to4}$	0.122	0.123	0.001	0.034	0.003	1.00
Transition prob. $P_{1\to5}$	0.119	0.120	0.001	0.028	0.004	1.00

Notes: Each replication follows the design in Table 4.1. Parameters are estimated by maximizing (4.9) using $S = 1024$ quasi–Monte Carlo (scrambled Sobol) draws and $|\mathcal{P}| = 20{,}000$ within-consumer pairs per replication. Panel B objects are computed by integrating simulated choice probabilities over an additional $S^{\text{eval}} = 256$ draws and averaging across choice occasions. "95% cov." reports two-sided 95% Wald coverage based on the inverse-Hessian approximation. Panel C shows inside choices only.

we evaluate the draw-specific probabilities $P_{ijt}^{[s]}(\theta)$ via logit smoothing with fixed scale $\omega_{\text{sm}} = 0.02$ (see the discussion around (3.19)). We initialize θ_δ at the multinomial logit estimates and set $(\lambda_P, \lambda_D) = (0.5, 0.5)$.

We report mean, bias, standard deviation, RMSE, and coverage for $(\alpha, \lambda_P, \lambda_D)$ as well as for several nonlinear objects implied by demand: the mean own-price elasticity, the mean cross-price elasticity, and the mean predicted choice probability of product 1. Elasticities and choice probabilities are computed by simulating choice probabilities under the estimated parameters using an additional set of $S^{\text{eval}} = 256$ draws and averaging across choice occasions.

4.5 Monte Carlo results

Table 4.2 summarizes the finite-sample performance of the pairwise likelihood estimator for the *AR-persistent model* in the baseline panel environment described above. Panel A reports results for the structural parameters $(\alpha, \lambda_P, \lambda_D)$. The estimator recovers the price coefficient accurately: the mean estimate of α is 0.979 against a true value of 1.000, implying a small bias of -0.021 with standard deviation 0.016 and RMSE 0.026. The persistent Cardell parameter λ_P is also estimated

precisely, with mean 0.337 relative to the true value 0.300 (bias 0.037; standard deviation 0.034). The autocorrelation parameter λ_D is less precisely estimated in this design: the mean estimate is 0.520 versus a true value of 0.600, with substantially larger sampling variability (standard deviation 0.180). Nonetheless, the estimates remain in a reasonable neighborhood of the truth and the reported Wald intervals are conservative here (coverage is 1.00 across parameters).

Panel B reports performance for implied demand objects computed from the estimated model. The mean own-price elasticity is close to its true benchmark (mean -3.938 vs. true -4.036), with modest bias and tight dispersion. The mean cross-price elasticity and the mean choice probability for product 1 are essentially unbiased (biases of -0.013 and 0.000, respectively). Two features are worth emphasizing when interpreting dispersion in Panel B: first, the "true" elasticities and choice probabilities vary across replications because they are evaluated at replication-specific simulated covariates and availability patterns; second, the RMSE therefore isolates how closely the estimator tracks the replication-specific benchmark, whereas the standard deviation of the estimates reflects both this estimation error and cross-replication variation in the underlying environment. Panel C shows that the repeat-choice rate and a set of transition probabilities out of product 1 exhibit negligible bias, indicating the estimation approach can recover these key objects.

Overall, the Monte Carlo evidence shows that the proposed AR–persistent specification can be estimated in panel settings using a computationally stable pairwise approach and can recover key demand objects while capturing meaningful intertemporal dependence.

5 Empirical application: OTC cough remedies

To illustrate the practical value of this approach, we apply our framework to the over-the-counter (OTC) cough remedies category using linked NielsenIQ scanner and household-panel data. Like many settings, consumer choices in this category exhibit substantial persistence, with households repeating their previous brand-formulation choice over half of purchase occasions. We compare our AR-persistent specification to standard random-coefficient logit benchmarks, including models with reduced-form lagged-choice state dependence.

5.1 Data

Our empirical application combines NielsenIQ Retail Measurement Services (RMS) scanner data with the NielsenIQ household panel (Homescan) over 2011–2013. RMS provides weekly UPC-level

sales and prices at the store level, while Homescan follows a panel of households and reports detailed shopping trips with UPC purchases and store identifiers. We study an over-the-counter (OTC) drug category—cough syrups and tablets—that features repeated purchases and meaningful brand differentiation. This provides an interesting setting to examine persistence in within-category choices.

To keep the choice problem tractable while retaining key substitution patterns, we aggregate UPCs to a moderate number of alternatives. Specifically, we pool RMS quantity sold to identify the top ten non-private-label brands, and we treat each *brand–formulation* pair among these leading brands as its own alternative. All remaining products, including private labels, are grouped into a single composite alternative, denoted $j = 0$ (*All Other*). Rather than define the outside option as not purchasing, we define the outside option in terms of this composite alternative. This provides a stable set of alternatives over time and provides a convenient within-category normalization for the mean utilities in the empirical analysis.[18]

We define a market as a store–week–year cell and index markets by $m \in \mathcal{M}$, where $m = (y, s, w)$ collects calendar year, store, and week-ending date. Alternatives are indexed by $j \in \{0, 1, \ldots, J\}$, where $j = 0$ denotes the composite "All Other" base alternative. Because package sizes vary substantially across the category, we standardize quantities to construct unit sales q_{jm} and compute within-category market shares as

$$\hat{s}_{jm} = \frac{q_{jm}}{\sum_{k=0}^{J} q_{km}}.$$

Here $\mathcal{J}_m \subseteq \{0, 1, \ldots, J\}$ denotes the set of alternatives treated as available in market m.[19]

Prices are constructed as standardized unit prices rather than package prices. We impute missing unit prices using average prices at the chain by Nielsen DMA level.[20] As a standard control for price endogeneity, we construct a Hausman-style instrument equal to the quantity-weighted mean unit price of the same alternative in all other areas during the same week.

The household panel data allow us to observe repeated purchase decisions and incorporate observable heterogeneity to discipline the demand model. We treat each shopping trip containing a category purchase as a purchase occasion. For household i, let $\tau = 1, \ldots, T_i$ index category purchase

[18] While much of the empirical literature defines the outside option as not purchasing, we argue this assumption is reasonable given our focus is on persistence in brand choice conditional on purchase, not on autocorrelation in whether a household purchases the category at all. Because category purchase is infrequent, the implied outside share would be close to one for most store-weeks and would mechanically dominate the variation in the data.

[19] Because scanner data can exhibit episodic sales and measurement noise, we assume an alternative is in the choice set in a store-week if it registers positive sales in at least two weeks within a 17-week centered window and appears in at least four weeks annually.

[20] If this is missing, we impute prices using averages at the Nielsen DMA level, then the national week-level averages, followed by linear temporal interpolation within the store-by-alternative series.

occasions, and let m denote the market.[21] The realized choice is denoted by $j_{i\tau} \in \{0, 1, \ldots, J\}$, and

$$y_{ij\tau} \equiv \mathbb{1}\{j_{i\tau} = j\}, \qquad j = 0, \ldots, J.$$

Each trip is linked to an RMS market by matching the store identifier and week-ending date, and the choice set is expanded over the carried alternatives in that market.

We form the micro estimation sample by cluster sampling households: we draw a subset of households without replacement from those with at least two category purchase occasions in the sample window, and we retain all eligible trips for each sampled household. We then ensure that the RMS store–week market set used on the macro side includes every market visited by these sampled trips.

We present summary statistics for the resulting macro and micro samples in Appendix Tables E.1 and E.2 respectively. The aggregate market-level data (Appendix Table E.1) consist of 278,159 store-week-alternative observations. The average standardized unit price across the sample is \$10.93. There is sizable variation in the Hausman instrument. The household-level estimation sample (Appendix Table E.2) encompasses 27,228 category purchase trips made by 14,611 distinct households. After applying our rolling carried-indicator screen, households face an average of 9.9 inside alternatives per trip. Appendix Table E.2 also details the demographic variation—such as household income, head age, and the presence of young children—that serves as the empirical basis for identifying observable heterogeneity in price sensitivities and brand preferences.

5.2 Demand specifications

We estimate four demand models that differ in how they capture persistence in household choices. The first two models are standard approaches in the applied literature, while the third and fourth models incorporate AR–persistent unobservables.

The sections above specialized to $\mu_{ijt} \equiv 0$ in order to simplify notation. We now reintroduce consumer-specific taste heterogeneity through a household-specific price coefficient. For household i on purchase occasion τ, let $m \equiv m_{i\tau}$. Utility is

$$u_{ij\tau} = \delta_{jm} + \mu_{ijm} + \epsilon_{ij\tau}.$$

[21] Technically, $m_{i\tau} \in \mathcal{M}$ denotes the market faced by household i on occasion τ since the market varies by household and occasion. We suppress subscripts for ease of notation.

We allow for heterogeneity in price sensitivity in the usual way by assuming

$$\mu_{ijm}(\vartheta) = \mu_{jm}(d_i; \vartheta) = -\alpha_i \, p_{jm}, \qquad \delta_{0m} \equiv 0, \tag{5.1}$$

where δ_{jm} is alternative-by-market mean utility, p_{jm} is price, and $\alpha_i > 0$ is a household-specific marginal disutility of price that is persistent over time. We parameterize

$$\alpha_i = \log\bigl(1 + \exp\bigl(\bar{\alpha} + d_i'\pi + \sigma \, z_i\bigr)\bigr), \qquad z_i \sim N(0,1), \tag{5.2}$$

so that price sensitivity varies with observed household shifters d_i and an unobserved component that is constant within household, while constraining the coefficient to be positive. In our AR-persistent specifications (Models 3 and 4) we set $\sigma = 0$ for tractability and focus on persistence in the unobservables.

As usual in demand applications, we specify the alternative–market mean utility as

$$\delta_{jm} = x_{jm}'\beta + \gamma_j + \xi_{jm},$$

where x_{jm} includes product-characteristic indicators (in practice, formulation measures) interacted with local market demographics, while γ_j denotes alternative fixed effects and ξ_{jm} is the remaining unobserved mean utility. Thus price enters separately through the household-specific term $-\alpha_i p_{jm}$, whereas δ_{jm} absorbs the usual observed and unobserved nonprice determinants of demand.

Model 1: Random-coefficient logit (persistent heterogeneity in price coefficient). We start by estimating a standard random-coefficient logit model. We assume i.i.d. extreme-value errors,

$$\epsilon_{ij\tau} = \varepsilon_{ij\tau}, \qquad \varepsilon_{ij\tau} \overset{iid}{\sim} \text{EV1},$$

so conditional on α_i the sequence $\{y_{i\tau}\}_\tau$ is independent over τ and choice probabilities take the standard form. Any persistence in observed brand choices is therefore driven by persistent heterogeneity in α_i and by persistence in observed covariates (prices and mean utilities).

Model 2: Random-coefficient logit with lagged-choice state dependence. We augment the utility index with an indicator for repeating the previous choice,

$$u_{ij\tau} = \delta_{jm} + \mu_{ijm} + \kappa \, \mathbb{1}\{j_{i,\tau-1} = j\} + \varepsilon_{ij\tau}, \quad \varepsilon_{ij\tau} \overset{iid}{\sim} \text{EV1}, \tag{5.3}$$

with the convention that $\mathbb{1}\{j_{i,\tau-1} = j\} = 0$ on the first observed occasion for each household. The parameter κ captures structural state dependence (habit persistence, switching costs, or inertia).[22] This is a standard approach in empirical work to capture persistence in choice in a reduced-form manner. Like Model 1, we still allow persistence in the random coefficient, however the presence of lagged choice in utility more directly allows for persistence in choice. However, the model does not allow serial correlation in $\epsilon_{ij\tau}$.

Model 3: AR-persistent unobservables. Model 3 imposes the recursive AR–persistent specification in Section 4.1 and sets $\kappa = 0$. The application-specific extension is that persistence in the transitory state depends on the elapsed time between purchase occasions. If $\Delta_{i\tau}$ denotes the number of weeks between household i's purchases $\tau - 1$ and τ, we set

$$\rho_{i\tau} = \lambda_D^{\Delta_{i\tau}}.$$

Hence repeat purchasing can arise either from stable household–product affinity or from transitory shocks whose influence decays more after longer gaps between purchases.

Model 4: AR-persistent unobservables with lagged-choice state dependence. Model 4 augments Model 3 with the lagged-choice term $\kappa \, \mathbb{1}\{j_{i,\tau-1} = j\}$. This lets the data distinguish persistence generated by the recursive AR–persistent unobservables from reduced-form state dependence. As before, variation in product availability across trips is informative because it shifts the lagged-choice regressor but not the latent AR–persistent component.

5.3 Endogenous prices and estimation: panel-CLEER

Our panel specification fits naturally within the standard micro-BLP framework treatment of endogenous prices (e.g., Petrin, 2002; Berry et al., 2004; Conlon and Gortmaker, 2025). For any

[22] Lagged-choice specifications can raise the classic initial-conditions concern that initial choices are correlated with persistent unobserved heterogeneity (Heckman, 1981). We use these models as reduced-form benchmarks. Our main results emphasize that allowing for AR–persistent unobservables reduces the estimated role of lagged-choice state dependence.

candidate value of the nonlinear parameters governing preference heterogeneity and persistence, the model delivers individual choice probabilities and implied market shares, which in turn pin down the vector of market-specific mean utilities through the usual inversion. We then estimate the remaining parameters using a panel analogue of the CLEER approach in Grieco et al. (2025), combining the micro likelihood with product-level exogeneity restrictions to address price endogeneity.

Macro shares and IV restrictions. For any candidate nonlinear parameter vector ϑ, let $\alpha(d;\vartheta)$ denote the implied price coefficient for household type d, and write

$$\mu_{jm}(d;\vartheta) \equiv -\alpha(d;\vartheta)\, p_{jm}.$$

The model-implied share in market m takes the mixed-logit form

$$s_{jm}(\delta,\vartheta) \equiv \mathbb{E}_d\left[\frac{a_{jm}\exp(\delta_{jm}+\mu_{jm}(d;\vartheta))}{\sum_{k=0}^{J} a_{km}\exp(\delta_{km}+\mu_{km}(d;\vartheta))}\right] \approx \frac{1}{R}\sum_{r=1}^{R}\frac{a_{jm}\exp(\delta_{jm}+\mu_{jm}(d_r;\vartheta))}{\sum_{k=0}^{J} a_{km}\exp(\delta_{km}+\mu_{km}(d_r;\vartheta))}. \tag{5.4}$$

The usual inversion for mean utilities therefore remains available.

Let N_{jm}^{rem} denote the aggregate counts used in the macro likelihood (either total counts or total counts net of the micro panel). The macro contribution is

$$\mathcal{L}_{\text{macro}}(\delta,\vartheta) = \sum_{m\in\mathcal{M}}\sum_{j=0}^{J} N_{jm}^{\text{rem}} \log s_{jm}(\delta,\vartheta). \tag{5.5}$$

To address price endogeneity, we follow Grieco et al. (2025) and combine this macro term with a product-level moment penalty based on excluded instruments, as we describe below.

Micro panel likelihood and persistence. The micro panel is indexed by within-household purchase occasions τ, ordered by purchase date. Let $\mathcal{P}_i$ denote the set of within-household occasion pairs used in the dependence correction. For each observed household–occasion pair (i,τ), let $\widehat{p}_{i\tau}(\delta,\vartheta)$ denote the simulated probability of the realized choice, and let $\widehat{p}_{i\tau\tau'}(\delta,\vartheta)$ denote the corresponding simulated pair probability. In Models 1–2 these objects are computed from standard logit probabilities averaged over the distribution of the household-specific price coefficient. In Models 3–4 they are computed by simulating the recursive AR–persistent specification in Section 4.1, allowing purchase-occasion decay $\rho_{i\tau} = \lambda_D^{\Delta_{i\tau}}$, and evaluating draw-specific probabilities.

The panel-CLEER micro criterion is

$$\mathcal{L}_{\text{micro}}(\delta,\vartheta) = \sum_i \sum_{\tau\in\mathcal{T}_i} \log \widehat{p}_{i\tau}(\delta,\vartheta) \;+\; \Omega_{\text{dep}}\, w \sum_i \sum_{(\tau,\tau')\in\mathcal{P}_i} \log\left(\frac{\widehat{p}_{i\tau\tau'}(\delta,\vartheta)}{\widehat{p}_{i\tau}(\delta,\vartheta)\,\widehat{p}_{i\tau'}(\delta,\vartheta)}\right), \tag{5.6}$$

where $w = N_{\text{cases}}/(2|\mathcal{P}|)$ is the same normalization used in (4.9). The first term is a singleton micro likelihood term. The second term is the dependence correction that uses within-household dependence in choices to identify (λ_P, λ_D). As discussed in Section 4.3, this correction remains comparatively tractable because it simulates only the low-dimensional latent components shared across two occasions rather than a full sequence likelihood. Appendix C provides implementation details and our baseline weighting scheme for Ω_{dep}.

Panel-CLEER estimator. Let ϑ collect the low-dimensional nonlinear parameters. In Models 1–2, ϑ contains the price-coefficient parameters (and κ in Model 2); in Models 3–4, ϑ contains the price-coefficient parameters together with (λ_P, λ_D) (and κ in Model 4). Let $\widetilde{Z}$ denote the matrix of excluded instruments after demeaning within product and standardization, and let $\widetilde{\delta}$ denote within-product deviations in mean utility. The product-level moments are $\widehat{m}(\delta) \equiv \widetilde{Z}'\widetilde{\delta}$ and the PLM penalty is $\chi_{\text{PLM}}(\delta) = \frac{1}{2}\,\widehat{m}(\delta)'W\,\widehat{m}(\delta)$, with W positive definite. The panel-CLEER estimator solves

$$(\widehat{\delta},\widehat{\vartheta}) \in \arg\max_{\delta,\vartheta} \left\{\mathcal{L}_{\text{macro}}(\delta,\vartheta) + \mathcal{L}_{\text{micro}}(\delta,\vartheta) - \tau_{PLM}\,\chi_{\text{PLM}}(\delta)\right\}. \tag{5.7}$$

with $\tau_{PLM} > 0$. Appendix C provides additional details, and Appendix C.5 describes standard errors for this composite-likelihood objective.

An advantage of building on the Grieco et al. (2025) approach is that it is a "conformant" estimator that attains the fastest feasible convergence rates given the identifying variation in the data. Our panel extension maintains a similar structure. Relative to Grieco et al. (2025), the macro term and PLM penalty are unchanged, while the panel extension only adds the dependence correction in (5.6).

5.4 Estimation results

Table 5.1 reports demand estimates for Models 1–4. Under the two benchmark specifications with i.i.d. extreme-value shocks (Models 1–2), the estimated dispersion in household price sensitivity is large. The implied distributions of household-specific price coefficients are very similar across these benchmarks, with $\mathbb{E}[\alpha_i] \approx 0.36$ and $\text{SD}[\alpha_i] \approx 0.87$ in both cases.

Table 5.1: Demand estimates under alternative persistence mechanisms

	Model 1 RC Logit		Model 2 RC Logit + lag		Model 3 AR-Persistent		Model 4 AR-Persistent + lag	
	Est	SE	Est	SE	Est	SE	Est	SE
Price-sensitivity index								
Price index mean ($\bar{\alpha}$)	-4.133	(0.871)	-4.057	(2.868)	-2.197	(0.281)	-2.195	(0.696)
Std. dev. of price coeff. (σ)	3.656	(0.750)	3.553	(3.323)				
Price sensitivity shifters								
Log income	0.916	(0.720)	0.884	(1.849)	0.202	(0.471)	0.199	(0.044)
HH head age	0.821	(0.427)	0.855	(0.705)	0.290	(0.089)	0.297	(0.037)
Household size	-0.871	(1.118)	-0.980	(2.422)	-0.153	(0.486)	-0.153	(0.281)
Has young child	0.656	(1.037)	0.737	(1.885)	0.009	(0.590)	0.015	(1.307)
2+ working adults	0.345	(0.356)	0.409	(0.749)	0.083	(0.738)	0.084	(0.290)
Persistence parameters								
Previous choice (κ)			3.280	(0.222)			0.236	(0.263)
Persistent parameter (λ_P)					0.124	(0.055)	0.129	(0.073)
AR decay parameter (λ_D)					0.512	(0.078)	0.473	(0.037)
Mean price coefficient ($\mathbb{E}[\alpha_i]$)	0.358		0.361		0.113		0.113	
Std price coefficient (SD[α_i])	0.865		0.861		0.043		0.043	
Objective: Macro	127,503.2		127,453.0		127,486.8		127,484.7	
Objective: Micro (singleton)	3,156.9		2,854.4		3,265.6		3,231.3	
Objective: Micro (dependence)	-214.0		-228.8		-698.6		-670.6	
Objective: PLM	3.286		3.590		0.010		0.007	
Objective: Total	130,449.4		130,082.2		130,053.7		130,045.4	

Notes: The table shows demand estimates for the OTC cough/cold category using linked NielsenIQ RMS (store–week) and Homescan (household-panel) data. The base alternative is "All Other" (mean utility normalized to zero); prices and instruments are relative to the base within market. Column (1): random-coefficient logit with persistent household price sensitivity. Column (2): Column (1) plus lagged-choice state dependence (κ). Column (3): AR-persistent unobservables per equation (4.2), estimated using a pairwise likelihood correction. Column (4): Column (3) plus lagged-choice state dependence. Standard errors in parentheses.

Model 2 augments the baseline random-coefficient logit with a lagged-choice indicator and estimates a very large state-dependence coefficient, $\kappa = 3.280$ (s.e. 0.222). Holding contemporaneous utilities fixed, this increases the odds of repeating the previous purchase by a factor of $\exp(\kappa) \approx 27$. At the same time, the price parameters in Model 2 remain close to those in Model 1, indicating that the lagged-choice term primarily absorbs persistence in observed sequences rather than changing contemporaneous substitution patterns.

Models 3–4 instead allow persistence through AR–persistent match-specific unobservables. In Model 3, the estimated dependence parameters $(\lambda_P, \lambda_D) = (0.124, 0.512)$ imply that repeat purchasing is generated by both a persistent household–product component and an innovation that

decays across purchase occasions. Relative to the i.i.d. benchmarks, introducing AR–persistent unobservables is accompanied by a substantially smaller mean price coefficient ($\mathbb{E}[\alpha_i] \approx 0.11$ rather than 0.36) and a much tighter implied distribution of price coefficients ($\mathrm{SD}[\alpha_i] \approx 0.04$ rather than 0.87), reflecting that less persistence needs to be rationalized through persistent price heterogeneity once match-level shocks are allowed.

Finally, Model 4 adds the lagged-choice regressor on top of the AR–persistent structure. The estimated coefficient on previous choice becomes small ($\kappa = 0.236$, implying an odds ratio of $\exp(\kappa) \approx 1.27$), and the remaining parameters—including the price coefficient and (λ_P, λ_D)—are similar to Model 3.

Taken together, these estimates suggest that, allowing for AR-persistent match shocks meaningfully reshapes the other estimated parameters relative to the benchmark models. Once persistent match quality and serially correlated taste shocks are accounted for, reduced-form state dependence (i.e., switching cost) plays little role in this setting. These results complement earlier work by Pakes et al. (2025), who use a moment inequality approach and also find that standard models overstate true state dependence in the context of health insurance.

5.5 Consumer-specific predictions and counterfactuals

An advantage of our approach is that it is straightforward to construct household-specific predictions. We use the estimates from Models 1–4 to construct household-specific posterior-predictive choice probabilities and evaluate how well each specification captures persistence in observed purchase sequences and predicts choices out of sample.

For each household, we construct weights over simulation draws proportional to the household's in-sample likelihood contribution. For the AR–persistent specifications (Models 3–4), these weights can be interpreted as an empirical-Bayes approximation to the posterior distribution of the latent Cardell components given the household's observed purchase history. We then form posterior-predictive choice probabilities and welfare objects for each subsequent trip by integrating draw-specific logit probabilities (and inclusive-value surplus expressions) with respect to these household-specific weights.

Repeat purchasing. Table 5.2 summarizes several measures of persistence in the household panel and their posterior-predictive analogs. In the data, households repeat their previous within-category choice on 58% of purchase occasions, and the fraction matching the product chosen two purchases

Table 5.2: Repeat purchasing and return after stockouts: observed vs. household-specific predicted probabilities

	Actual	Model 1 RC Logit	Model 2 RC Logit + lag	Model 3 AR-Persistent	Model 4 AR-Persistent + lag
Repeat choice fraction	0.581	0.152	0.558	0.550	0.567
Repeat choice fraction ($t = t-2$)	0.555	0.145	0.375	0.537	0.531
Repeat choice fraction w/ stockout ($t = t-2$)	0.579	0.126	0.234	0.513	0.496

Notes: "Actual" reports the fraction of trips for which the household chooses the same alternative as on its kth previous category purchase, $\mathbb{E}[1\{y_{it} = y_{i,t-k}\}]$. Model columns report the mean posterior-predictive probability assigned to the lagged choice, $\mathbb{E}[p_{it}(y_{i,t-k})]$, where posterior weights are constructed from the household's in-sample purchase history. "Stockout" conditions on the product chosen at $t-2$ being unavailable at $t-1$ and available again at t.

earlier is nearly as high (55%).

The baseline random-coefficient logit (Model 1), where unobserved shocks are i.i.d. over time and persistence can only arise through persistent heterogeneity and observed covariates, dramatically understates this persistence: it assigns only 0.15 probability to repeating the last choice and 0.14 to matching the choice two purchases earlier. Adding a lagged-choice indicator (Model 2) increases the one-lag repeat probability to 0.56. While this is close to the actual probability, it improves the two-lag repeat probability only modestly (to 0.38) which is still far from the actual value.

In contrast, the AR–persistent specifications (Models 3–4) come much closer to the data. Model 3 predicts repeat probabilities of 0.55 (one-lag) and 0.54 (two-lag), and Model 4 delivers very similar predictions (0.57 and 0.53). The similarity between Models 3 and 4 is consistent with the small estimated lagged-choice coefficient in Model 4 (Table 5.1): once persistent match quality and serially correlated shocks are present, adding an additional last-choice "bump" does not materially change the predicted persistence moments.

The last row in Table 5.2 focuses on "return after stockout": conditioning on the product purchased at $t-2$ being unavailable at $t-1$ but available again at t, households return to the earlier product with probability 0.579 in our sample. Both i.i.d. benchmarks substantially underpredict this behavior (0.13 in Model 1 and 0.23 in Model 2). The lagged-choice specification does not help much for this moment because the forced switch at $t-1$ mechanically resets the lag indicator. The AR–persistent model does substantially better, predicting a return probability of 0.51, and again Model 4 is similar.

We next evaluate how these persistence mechanisms translate into out-of-sample predictive performance.

Table 5.3: Household-specific out-of-sample fraction correctly predicted by forecast horizon

	Forecast horizon for held out sample					
Model	1	2	3	4	5	Avg.
Panel A: Fraction correctly predicted						
Random-coefficient logit	0.474	0.488	0.504	0.446	0.405	0.463
Random-coefficient logit + prev choice	0.584	0.466	0.408	0.353	0.322	0.426
AR–persistent	0.507	0.537	0.507	0.545	0.457	0.511
AR–persistent + prev choice	0.507	0.529	0.496	0.540	0.452	0.505
Panel B: Mean predictive probability of realized choice						
Random-coefficient logit	0.284	0.298	0.298	0.278	0.255	0.283
Random-coefficient logit + prev choice	0.371	0.305	0.274	0.250	0.225	0.285
AR–persistent	0.402	0.421	0.400	0.418	0.346	0.397
AR–persistent + prev choice	0.399	0.412	0.392	0.419	0.344	0.393

Notes: The table reports trip-level out-of-sample predictive performance for the first $h = 1, \ldots, H$ out-of-sample trips after each household's last in-sample purchase ("horizon" is the index after the last in-sample trip). Predictions are household-specific posterior-predictive objects obtained by integrating over simulation draws using empirical-Bayes (posterior) weights constructed from in-sample choices. Panel A reports the fraction of trips for which the modal predicted alternative equals the realized choice. Panel B reports the mean predictive probability of realized choice. The "Avg." column averages over all out-of-sample trips within the horizons shown.

Out-of-sample prediction. Table 5.3 reports trip-level out-of-sample performance by forecast horizon, using household-specific posterior-predictive probabilities. The lagged-choice specification (Model 2) performs best at the one-step-ahead horizon: the fraction correctly predicted is 0.584 at horizon 1, compared with 0.474 in the baseline random-coefficient logit (Model 1). However, its performance deteriorates quickly as the forecast horizon increases (falling to 0.322 by horizon 5), and its average over the five horizons is lower than the baseline (0.426 versus 0.463). This pattern is consistent with the lagged-choice indicator capturing very short-run repetition but providing little guidance once several purchase occasions have elapsed.

By contrast, the AR–persistent model delivers consistently strong performance across horizons. It attains the highest average fraction correctly predicted (0.511) and assigns substantially higher probability to the realized choice on held-out trips (Panel B average 0.397, versus about 0.28 in Models 1–2). Model 4 again tracks Model 3 closely. Overall, allowing for persistent match quality and serially correlated shocks improves household-level forecasting in a way that remains robust at longer horizons, while adding an additional lagged-choice term on top of this structure yields little incremental predictive gain.

Table 5.4: Change in consumer surplus from removing the highest-share product

Model	1	2	3	4	5
RC logit	-4.045	-4.252	-3.900	-3.905	-3.340
RC logit + prev choice	-9.766	-9.694	-8.936	-8.207	-7.330
AR–persistent	-1.123	-1.130	-1.158	-1.129	-0.942
AR–persistent + prev choice	-1.138	-1.158	-1.265	-1.172	-0.971

Notes: The counterfactual removes the inside alternative with the highest observed in-sample share from the choice set, holding prices and mean utilities fixed (partial equilibrium). Each cell reports the trip-weighted mean change in expected consumer surplus, $\mathbb{E}[\Delta CS_{it}]$, among out-of-sample trips at horizon h after the last in-sample purchase ($h = 1, 2, \dots$). Consumer surplus is computed from the log-sum inclusive value and converted to monetary units using the household-specific marginal disutility of price; in the AR–persistent model the inclusive value is additionally scaled by the implied logit scale factor (see Appendix D). Predictions and welfare objects are household-specific posterior-predictions constructed from in-sample choices. Negative values indicate welfare losses from product removal.

Welfare from product removal. Table 5.4 reports the mean per-trip change in expected consumer surplus from a partial-equilibrium counterfactual that removes the highest-share inside product from the choice set, evaluated on the first five held-out purchase occasions after each household's last in-sample trip. Across all specifications and horizons, removing the category leader reduces consumer surplus, but the implied magnitude is highly sensitive to how persistence is modeled.

Under the baseline random-coefficient logit (Model 1), the mean welfare loss is on the order of -3.3 to -4.3 per trip across horizons. Introducing a lagged-choice term (Model 2) roughly doubles these losses: the mean loss is -9.77 at horizon 1 and remains around -7 to -9 through horizon 5. In contrast, the AR–persistent model (Model 3) implies much smaller losses, around -1.1 per trip across horizons. Adding the lagged-choice regressor on top of the AR–persistent structure (Model 4) has only a minor effect: the implied losses are very close to Model 3 at each horizon.

These patterns highlight that welfare counterfactuals involving product availability depend not only on contemporaneous substitution patterns, but also on how the model attributes repeat purchasing—to a last-choice utility "bump" versus persistent match quality and serially correlated shocks that are not mechanically tied to the most recent observed choice. In this setting, once AR–persistent unobservables are allowed, the additional lagged-choice term is small and does not meaningfully change predicted behavior or welfare.

6 Conclusion

Discrete-choice demand models are widely used in a variety of settings in which persistence is a first-order issue, yet flexibly modeling serially correlated unobservable preferences in long panels

has historically been hindered by the computational burden of high-dimensional integration. In this paper, we develop a tractable panel discrete-choice framework that introduces correlated unobserved utility across products and over time via simulated Cardell variance components. By exploiting the convolution property of the Cardell distribution, we simulate low-dimensional latent components while integrating out the remaining extreme-value term analytically. This approach yields closed-form logit choice probabilities conditional on the simulated latent components, accommodating complex structures like overlapping groupings and serial dependence.

The framework naturally nests standard static models, such as single and multi-level nested logit, while providing a foundation for more flexible specifications. We specifically highlight a parsimonious AR-persistent specification that separates stable match-specific quality from autoregressive taste shocks that decay over time. Estimation in long panels is feasible when the model is estimated using a pairwise composite likelihood. Our Monte Carlo experiments confirm that this estimator accurately recovers structural parameters, produces reliable elasticity estimates.

We illustrate the empirical relevance of the framework in the over-the-counter cough-remedies category using linked NielsenIQ household-panel and scanner data. To combine flexible micro-level unobserved heterogeneity with aggregate share information, we use a panel adaptation of the CLEER estimator. In this setting, the AR-persistent specification provides a substantially better account of observed behavior than standard random-coefficients logit and lagged-choice models. In particular, it more accurately matches repeat purchasing and return-after-stockout patterns, and it avoids attributing persistent behavior to spurious state dependence or excessive heterogeneity in price sensitivity. These differences matter quantitatively. The AR-persistent model delivers better out-of-sample predictions and implies materially different substitution and welfare patterns in counterfactual exercises.

More broadly, the paper shows that richer serial dependence in demand can be incorporated without sacrificing the tractability that makes logit-based models useful in practice. While multinomial probit models with Gaussian errors offer extensive flexibility through a largely unrestricted covariance structure, that flexibility typically comes at the cost of evaluating high-dimensional multivariate normal probabilities, which becomes burdensome in long panels and large choice sets. Our approach instead uses simulated Cardell variance components to accommodate economically interpretable dependence patterns while preserving exact logit marginals conditional on simulated dependence terms. This is useful not only for the applications studied here, but also for a wider class of demand settings in which choices reflect durable attachments, evolving match quality, or

repeated exposure to product-specific shocks. We therefore view the approach developed here as a practical foundation for future work on richer panel demand models with correlated preferences across products and over time.

References

Berry, S., J. Levinsohn, and A. Pakes (1995, July). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841.

Berry, S., J. Levinsohn, and A. Pakes (2004, February). Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy 112*(1), 68–105.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Berry, S. T. and P. A. Haile (2021, September). Foundations of Demand Estimation.

Bertschek, I. and M. Lechner (1998, December). Convenient estimators for the panel probit model. *Journal of Econometrics 87*(2), 329–371.

Bhat, C. R. (2014, July). The Composite Marginal Likelihood (CML) Inference Approach with Applications to Discrete and Mixed Dependent Variable Model. *Foundations and Trends in Econometrics 7*(1), 1–117.

Blevins, J. R. (2016, August). Sequential Monte Carlo Methods for Estimating Dynamic Microeconomic Models. *Journal of Applied Econometrics 31*(5), 773–804.

Boehm, E. (2025). Intermediation, Choice Frictions, and Adverse Selection: Evidence from the Chilean Pension Market.

Brown, Z. Y., M. L. Egan, J. Jeon, C. Jin, and A. A. Wu (2026, January). Why Do Index Funds Have Market Power? Quantifying Frictions in the Index Fund Market.

Brown, Z. Y. and J. Jeon (2024). Endogenous Information and Simplifying Insurance Choice. *Econometrica 92*(3), 881–911.

Cantillo, V., J. D. D. Ortúzar, and H. C. W. L. Williams (2007, May). Modeling Discrete Choices in the Presence of Inertia and Serial Correlation. *Transportation Science 41*(2), 195–205.

Cardell, N. S. (1997). Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity. *Econometric Theory 13*(2), 185–213.

Chambers, J. M., C. L. Mallows, and B. W. Stuck (1976). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association 71*(354), 340–344.

Chintagunta, P., E. Kyriazidou, and J. Perktold (2001, July). Panel data analysis of household brand choices. *Journal of Econometrics 103*(1-2), 111–153.

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

Conlon, C. and J. Gortmaker (2025, January). Incorporating Micro Data into Differentiated Products Demand Estimation with PyBLP. *Journal of Econometrics*, 105926.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010, August). State dependence and alternative explanations for consumer inertia. *The RAND Journal of Economics 41*(3), 417–445.

Fernández-Villaverde, J. and J. F. Rubio-Ramírez (2007, October). Estimating Macroeconomic Models: A Likelihood Approach. *Review of Economic Studies 74*(4), 1059–1087.

Galichon, A. (2022, April). On the Representation of the Nested Logit Model. *Econometric Theory 38*(2), 370–380.

Galichon, A. and B. Salanié (2022, September). Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *The Review of Economic Studies 89*(5), 2600–2629.

Godambe, V. P. (1960). An Optimum Property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics 31*(4), 1208–1211.

Grieco, P. L., C. Murry, J. Pinkse, and S. Sagl (2025, January). Optimal Estimation of Discrete Choice Demand Models with Consumer and Product Data. Technical Report w33397, National Bureau of Economic Research, Cambridge, MA.

Heckman, J. J. (1981). Statistical models for discrete panel data. *Structural analysis of discrete data with econometric applications 114*, 178.

Honore, B. E. and E. Kyriazidou (2000, July). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica 68*(4), 839–874.

Huber, P. (1967). The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press 1*, 221–233.

Kanter, M. (1975, August). Stable Densities Under Change of Scale and Total Variation Inequalities. *The Annals of Probability 3*(4), 697–707.

Kaye, A. P. (2024). The Personalization Paradox: Welfare Effects of Personalized Recommendations in Two-Sided Digital Markets. *Working Paper*.

Keane, M. (2015, May). Panel Data Discrete Choice Models of Consumer Demand. In B. H. Baltagi (Ed.), *The Oxford Handbook of Panel Data* (1 ed.)., pp. 548–582. Oxford University Press.

Keane, M. P. (1994, January). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica 62*(1), 95.

Keane, M. P. (1997). Modeling Heterogeneity and State Dependence in Consumer Choice Behavior. *Journal of Business & Economic Statistics 15*(3), 310–327.

Lachaab, M., A. Ansari, K. Jedidi, and A. Trabelsi (2006, March). Modeling preference evolution in discrete choice models: A Bayesian state-space approach. *Quantitative Marketing and Economics 4*(1), 57–81.

McFadden, D. (1973). McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior. In Zarembka, P., Ed., Frontiers in Econometrics, Academic Press, 105-142. - References - Scientific Research Publishing.

McFadden, D. (1978). Modelling the Choice of Residential Location.

Norets, A. (2009). Inference in Dynamic Discrete Choice Models With Serially Correlated Unobserved State Variables. *Econometrica 77*(5), 1665–1682.

Pakes, A. and J. Porter (2024). Moment inequalities for multinomial choice with fixed effects. *Quantitative Economics 15*(1), 1–25.

Pakes, A., J. R. Porter, M. Shepard, and S. Calder-Wang (2025). Unobserved Heterogeneity, State Dependence, and Health Plan Choices. *NBER*.

Petrin, A. (2002, August). Quantifying the Benefits of New Products: The Case of the Minivan. *Journal of Political Economy 110*(4), 705–729.

Seetharaman, P. B. (2004, May). Modeling Multiple Sources of State Dependence in Random Utility Models: A Distributed Lag Approach. *Marketing Science 23*(2), 263–271.

Shi, X., M. Shum, and W. Song (2018). Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity. *Econometrica 86*(2), 737–761.

Tuzcuoglu, K. (2023, April). Composite Likelihood Estimation of an Autoregressive Panel Ordered Probit Model with Random Effects. *Journal of Business & Economic Statistics 41*(2), 593–607.

Varin, C., N. Reid, and D. Firth (2011). An Overview of Composite Likelihood Methods. *Statistica Sinica 21*(1), 5–42.

Vovsha, P. (1997, January). Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record: Journal of the Transportation Research Board 1607*(1), 6–15.

Wen, C.-H. and F. S. Koppelman (2001, August). The generalized nested logit model. *Transportation Research Part B: Methodological 35*(7), 627–641.

Appendix

A Variance component layering and Cardell primitives

This appendix section provides details about the layering approach for the Cardell errors and practical strategies for constructing these layered representations that allow for flexible substitution patterns.

Layers and nests. Let $\ell = 1, \ldots, K$ index ordered *layers*. Each layer ℓ defines a partition of the alternatives $\mathcal{J}$ into disjoint nests $\mathcal{G}_\ell = \{1, \ldots, G_\ell\}$ and a mapping $g_\ell : \mathcal{J} \to \mathcal{G}_\ell$ that assigns each product j to its unique nest $g_\ell(j)$ in that layer. Across layers, nests may overlap: the same product can belong to different nests in different layers. Within a layer, nests are disjoint by construction.

It is sometimes convenient to index a *variance component* by $k = (\ell, g)$, i.e., the gth nest in layer ℓ. We write $\lambda_{\ell g} \in (0, 1]$ for the Cardell (layer) parameter associated with nest g in layer ℓ.

Layer minimum and common scale. Define the layerwise minimum parameter

$$\underline{\lambda}_\ell \;\equiv\; \min_{g \in \mathcal{G}_\ell} \lambda_{\ell g}.$$

Layering enforces a common EV1 scale across alternatives by ensuring that *every* alternative receives a layer-ℓ contribution with the same Cardell scale $\underline{\lambda}_\ell$. When $\lambda_{\ell g} > \underline{\lambda}_\ell$ for some nests, we achieve this by adding an adjustment primitive.

Adjusted Cardell primitives. For each consumer i, nest $g \in \mathcal{G}_\ell$, and (when relevant) time t, draw mutually independent primitives

$$\nu_{igt}^{(\ell)} \sim C(\lambda_{\ell g}) \qquad \text{and} \qquad \nu_{igt}^{\text{adj},(\ell)} \sim C\left(\frac{\underline{\lambda}_\ell}{\lambda_{\ell g}}\right),$$

with the understanding that persistent layers suppress the time index t. Define the adjusted Cardell shock

$$\widetilde{\nu}_{igt}^{(\ell)} \;\equiv\; \nu_{igt}^{(\ell)} + \lambda_{\ell g}\, \nu_{igt}^{\text{adj},(\ell)}.$$

By the Cardell recursion, $\widetilde{\nu}_{igt}^{(\ell)} \sim C(\underline{\lambda}_\ell)$ for every g. If $\lambda_{\ell g} = \underline{\lambda}_\ell$, then $\nu_{igt}^{\text{adj},(\ell)} \sim C(1)$ is degenerate at zero, so $\widetilde{\nu}_{igt}^{(\ell)} = \nu_{igt}^{(\ell)}$.

Dynamic lags and scale recursion. Each layer ℓ is assigned a maximum lag length $L_\ell \in \{0, 1, 2, \dots\}$, shared by all nests in that layer. For layers used only in persistent or within-period components, set $L_\ell = 0$. For layers used in dynamic components, L_ℓ can be positive. Define the residual scale passed forward across layers by

$$\Lambda_0 \equiv 1, \qquad \Lambda_\ell \equiv \Lambda_{\ell-1}\, \underline{\lambda}_\ell^{1+L_\ell}, \qquad \ell = 1, \dots, \mathcal{L}.$$

The common residual EV1 scale in (2.7) is $\Lambda \equiv \Lambda_K$.

Layer contribution. For any (i, j, t), define the contribution of layer ℓ as

$$\eta_{ijt}^{(\ell)} \;\equiv\; \Lambda_{\ell-1} \sum_{s=0}^{L_\ell} \underline{\lambda}_\ell^{s}\, \widetilde{\nu}_{i,g_\ell(j),t-s}^{(\ell)}. \tag{A.1}$$

This expression induces serial dependence when $L_\ell > 0$ because the same primitive $\widetilde{\nu}_{i,g,t}^{(\ell)}$ enters utilities at multiple dates via $t - s$.

Allocating layers to persistent, dynamic, and within-period components. Let $\mathcal{K}_{\zeta^P}, \mathcal{K}_{\tau^P}, \mathcal{K}_{\zeta^D}, \mathcal{K}_{\tau^D}, \mathcal{K}_{\tau^0} \subseteq \{1, \dots, K\}$ denote the subsets of layers feeding each term in (2.8). These sets are typically taken to be disjoint and to form a partition of $\{1, \dots, K\}$.

Using $\eta_{ijt}^{(\ell)}$ above, define

$$\begin{aligned}
\zeta_{ij}^{P} &\equiv \sum_{\ell \in \mathcal{L}_{\zeta^P}} \eta_{ijt}^{(\ell)} && (L_\ell = 0,\ \text{no } t \text{ dependence}),\\
\tau_{i,g(j)}^{P} &\equiv \sum_{\ell \in \mathcal{L}_{\tau^P}} \eta_{ijt}^{(\ell)} && (L_\ell = 0,\ \text{no } t \text{ dependence}),\\
\zeta_{ijt}^{D} &\equiv \sum_{\ell \in \mathcal{L}_{\zeta^D}} \eta_{ijt}^{(\ell)} && (L_\ell \geq 0,\ \text{lags } t - s),\\
\tau_{i,g(j),t}^{D} &\equiv \sum_{\ell \in \mathcal{L}_{\tau^D}} \eta_{ijt}^{(\ell)} && (L_\ell \geq 0,\ \text{lags } t - s),\\
\tau_{i,g(j),t}^{0} &\equiv \sum_{\ell \in \mathcal{L}_{\tau^0}} \eta_{ijt}^{(\ell)} && (L_\ell = 0,\ \text{within-period only}).
\end{aligned}$$

Finally, the three blocks in (2.8) are

$$\eta_{ij}^{P} \equiv \zeta_{ij}^{P} + \tau_{i,g(j)}^{P}, \qquad \eta_{ijt}^{D} \equiv \zeta_{ijt}^{D} + \tau_{i,g(j),t}^{D}, \qquad \eta_{ijt}^{0} \equiv \tau_{i,g(j),t}^{0}.$$

Remark (EV1 normalization). Under the mutual-independence assumptions on the primitives across distinct indices and the Cardell conjugacy recursion, the layered sum $\eta_{ijt} = \sum_{\ell=1}^{\mathcal{L}} \eta_{ijt}^{(\ell)}$ is marginally distributed as $C(\Lambda)$. Consequently, $\epsilon_{ijt} = \eta_{ijt} + \Lambda\varepsilon_{ijt}$ is marginally standard EV1, as stated in (2.7).

A.1 Layer ordering and continuation scales

Cardell conjugacy and continuation scales. Given that Cardell conjugacy implies $\nu+\lambda\varepsilon \sim EV1$ and yields the closure recursion used in the layered construction, this allows us to introduce structured dependence through η_{ijt} while preserving an extreme-value marginal for $\epsilon_{ijt} = \eta_{ijt} + \Lambda\varepsilon_{ijt}$.

Layering implements this recursively. With residual-scale recursion (illustrated in Appendix A) the remaining EV1 scale after processing layers $1, \ldots, \ell$ is

$$\Lambda_0 \equiv 1, \qquad \Lambda_\ell \equiv \Lambda_{\ell-1}\, \lambda_\ell^{1+L_\ell}, \qquad \Lambda \equiv \Lambda_L, \tag{A.2}$$

and each layer contributes $\eta_{ijt}^{(\ell)}$ scaled by $\Lambda_{\ell-1}$. A useful closure implication is the *continuation property*: for any cutoff layer $\bar{\ell}$, the tail

$$R_{ijt}^{(\bar{\ell})} \;\equiv\; \sum_{\ell=\bar{\ell}}^{L} \eta_{ijt}^{(\ell)} + \Lambda\varepsilon_{ijt} \tag{A.3}$$

is again extreme value, with continuation scale given by the recursion after layers $1, \ldots, \bar{\ell}-1$ (i.e., $\Lambda_{\bar{\ell}-1}$). Importantly, $R_{ijt}^{(\bar{\ell})}$ need not be i.i.d. across j if the tail still contains shared (nest/group) components, but the EV1 closure is what enables analytic integration of *blocks* of lower-layer primitives once we condition on a reduced set of simulated components.

Why ordering matters: which layers should be "high" vs "low"? Layering is organizational: each layer is a partition of products into disjoint nests, and products can "load" on multiple (possibly overlapping) groupings by appearing in multiple layers (e.g., brand in one layer, category in another). The practical reason to be deliberate about the ordering is computational: the continuation property only becomes useful for objective construction if the set of layers you want to integrate out corresponds to a *tail* of the ordered list.

A simple rule of thumb is:

- **Put within-occasion dependence layers early (high layers).** These are the layers that

create contemporaneous correlation across alternatives inside a choice set (e.g., within-period nest–time components). Because these terms couple utilities across j at time t, they are typically the terms that *force* simulation or numerical integration in per-occasion probabilities. Putting them early keeps them "exposed" in any partial-integration scheme and, because layering maintains a common continuation scale across alternatives, the remaining tail still slots into a well-defined logit choice probability after conditioning.

- **Put i.i.d.-within-occasion layers late (low layers).** These include alternative–time components that are independent across j within a given (i, t) choice situation—*even if they are persistent or serially correlated over t for a given product.* When such terms sit in the low layers, they can often be absorbed into the EV1 "tail" and integrated out analytically (or, in some special cases, disappear entirely from marginal per-occasion choice probabilities), reducing how much simulation is needed for the objective.

- **Put dynamic layers with long lags in the lowest layer.** Dynamic layers with multiple lags tend to multiply the continuation scale many times, driving their associated Λ_ℓ close to zero. This is not problematic so long as the dynamic component is placed in the final layer immediately above the marginal i.i.d. term $\Lambda\varepsilon$, so that any near-degeneracy is confined to the last continuation scale and does not distort the scaling of higher layers.

A.2 Layering with Truncated AR(1) Cardell shocks

The motivation for (**??**) is that it preserves the EV1 marginal while generating serial dependence. To see this, define a sequence of intermediate shocks constructed from the oldest to the newest innovation:

$$W_{ijt}^{(0)} \equiv \varepsilon_{ijt}, \qquad W_{ijt}^{(m)} \equiv \nu_{ij,t-L+m} + \lambda_D \, W_{ijt}^{(m-1)}, \quad m = 1, \ldots, L. \tag{A.4}$$

Because each $\nu_{ij,t-L+m} \sim C(\lambda_D)$ is independent of $W_{ijt}^{(m-1)}$ and $W_{ijt}^{(m-1)} \sim$ EV1 by induction, the Cardell property implies $W_{ijt}^{(m)} \sim$ EV1 for all m. Expanding $W_{ijt}^{(L)}$ yields exactly (**??**), so $\epsilon_{ijt} \sim$ EV1 for each (i, j, t) even though the same innovation $\nu_{ij,t}$ enters multiple periods and generates intertemporal dependence.

B Special Case Details

B.1 Finite-memory trip representation.

In the finite-memory variant, the memory window $[t_k - L,\, t_k]$ may not align with observed trip times, so the window boundary can fall in the interior of an inter-trip interval. Let $s^*(k)$ denote the trip index at which this occurs:

$$t_{s^*-1} < t_k - L \leq t_{s^*}.$$

Blocks for trips $s > s^*$ fall entirely within the memory window and are retained in full. The block for trip s^* is split at the boundary: only the primitives arriving after $t_k - L$ are retained, giving a truncated accumulation $\zeta^D_{ij,(t_k-L,\,t_{s^*}]} \sim C\big(\lambda_D^{t_{s^*}-t_k+L}\big)$. Blocks for trips $s < s^*$ fall entirely outside the window and are absorbed into the residual EV1 term. The resulting expression is

$$\epsilon_{ij,t_k} \;=\; \sum_{s=s^*+1}^{k} \lambda_D^{t_k-t_s}\, \zeta^D_{ij,(t_{s-1},\,t_s]} \;+\; \lambda_D^{t_k-t_{s^*}}\, \zeta^D_{ij,(t_k-L,\,t_{s^*}]} \;+\; \lambda_D^{L}\, \varepsilon_{ij,t_k}. \tag{B.1}$$

When $t_k - L$ coincides with a trip time t_{s^*}, the truncated block vanishes and the expression reduces to a sum over full blocks from $s^* + 1$ onward.

C Panel-CLEER estimation details

This appendix provides implementation details for the panel-CLEER estimator summarized in Section 5.3. We follow Grieco et al. (2025) in (i) using an aggregate-share likelihood in the high-dimensional mean utilities δ and (ii) enforcing instrumental-variables restrictions through a product-level moment (PLM) penalty. The extension specific to this paper is the treatment of *panel* micro data under the AR–persistent specifications: we exploit market-level choice probabilities on the macro side and append a pairwise (composite-likelihood) correction that uses within-household dependence to identify persistence parameters.

C.1 Data structure, indices, and market-level choice probabilities

Let $m \in \mathcal{M}$ index markets, where a market is a store–week–year cell, and let $j \in \{0, 1, \ldots, J\}$ index alternatives, with the composite base alternative $j = 0$ normalized to $\delta_{0m} \equiv 0$ and $p_{0m} \equiv 0$. Let $a_{jm} \in \{0, 1\}$ denote availability in market m.

On the household side, let $\tau = 1, \ldots, T_i$ index household i's category purchase occasions after sorting trips by purchase date, and let $m(i,\tau) \in \mathcal{M}$ denote the market visited on occasion τ.

For the empirical specification, utility is

$$u_{ij\tau} = \delta_{j,m(i,\tau)} + \mu_{ij,m(i,\tau)} + \epsilon_{ij\tau}, \qquad \mu_{ijm} = -\alpha_i\, p_{jm}, \qquad \alpha_i = \log\bigl(1 + \exp\bigl(\bar{\alpha} + d_i'\pi\bigr)\bigr)\,,$$

with $\sigma = 0$ so the price coefficient varies only with observed household characteristics. In Model 4, the lagged-choice term $\kappa\, \mathbb{1}\{y_{i,\tau-1} = j\}$ is added on the micro side; it does not enter the macro shares.

Let ϑ collect the low-dimensional nonlinear parameters. For the macro component, ϑ enters through the price-coefficient parameters, and we write

$$\mu_{jm}(d;\vartheta) \equiv -\alpha(d;\vartheta)\, p_{jm}.$$

The model-implied market share is therefore

$$s_{jm}(\delta,\vartheta) \equiv \mathbb{E}_d\left[\frac{a_{jm}\exp(\delta_{jm} + \mu_{jm}(d;\vartheta))}{\sum_{k=0}^{J} a_{km}\exp(\delta_{km} + \mu_{km}(d;\vartheta))}\right] \approx \frac{1}{R}\sum_{r=1}^{R}\frac{a_{jm}\exp(\delta_{jm} + \mu_{jm}(d_r;\vartheta))}{\sum_{k=0}^{J} a_{km}\exp(\delta_{km} + \mu_{km}(d_r;\vartheta))}. \tag{C.1}$$

Serial dependence in household unobservables affects *joint* choice probabilities across purchase occasions, but it does not alter (C.1): the macro component depends only on market-level marginals.

C.2 Macro likelihood from aggregate shares and micro–macro accounting

Let N_{jm}^{tot} denote aggregate counts for alternative j in market m (or the implied counts from scanner shares and category size). Define

$$N_{jm}^{\text{micro}} \equiv \sum_{i}\sum_{\tau\in\mathcal{T}_i} \mathbb{1}\{m_{i\tau} = m\}\; y_{ij\tau}, \qquad N_{jm}^{\text{rem}} \equiv N_{jm}^{\text{tot}} - N_{jm}^{\text{micro}}.$$

If one instead treats the macro and micro samples as disjoint, set $N_{jm}^{\text{rem}} = N_{jm}^{\text{tot}}$.[23] The macro contribution is

$$\mathcal{L}_{\text{macro}}(\delta,\vartheta) = \sum_{m\in\mathcal{M}}\sum_{j=0}^{J} N_{jm}^{\text{rem}} \log s_{jm}(\delta,\vartheta). \tag{C.2}$$

[23] In the application, NielsenIQ Homescan purchases are not a strict subset of RMS store transactions. In our baseline implementation we subtract micro counts as an approximate "no double counting" correction, and we verify that results are similar when setting $N_{jm}^{\text{rem}} = N_{jm}^{\text{tot}}$.

Together with the singleton part of the micro criterion below, (C.2) reconstructs the population's cross-sectional likelihood: the micro panel contributes the observed households, while the aggregate shares contribute the remainder (Grieco et al., 2025). The remaining identifying content of the panel for persistence parameters comes from within-household dependence over purchase occasions.

C.3 Micro panel likelihood: simulation and pairwise sequence correction

The micro data are a panel of households $i \in \mathcal{I}$ observed on purchase occasions $\tau \in \mathcal{T}_i$, with realized choices $j_{i\tau}$ and markets $m_{i\tau}$. Because unobservables are serially dependent, the exact sequence likelihood is high-dimensional and does not factorize over τ. We therefore use the pairwise correction described in Section 4.3, evaluated using simulated probabilities under the AR–persistent specification.

For simulation draw $s = 1, \ldots, S$, simulate $\eta_{ij}^{[s]}$ and the recursive state $\varepsilon_{ij\tau}^{\star,[s]}$ from Section 4.1. In the NielsenIQ implementation, the purchase-occasion persistence parameter is $\rho_{i\tau} = \lambda_D^{\Delta_{i\tau}}$, where $\Delta_{i\tau}$ is the number of weeks between purchases. The draw-specific latent component entering utility is

$$\nu_{ij\tau}^{[s]} = \eta_{ij}^{[s]} + \lambda_P \varepsilon_{ij\tau}^{\star,[s]}.$$

Let

$$\bar{u}_{ij\tau}(\delta, \vartheta) \equiv \delta_{jm} + \mu_{ijm} + \kappa \, \mathbb{1}\{j_{i,\tau-1} = j\}, \qquad m \equiv m_{i\tau},$$

with the convention that $\kappa = 0$ in Model 3 and that the lagged-choice indicator is zero on the first observed occasion for each household and for the base alternative.

Because the recursive AR specification does not yield a closed-form conditional on shocks, we evaluate draw-specific choice probabilities using logit-smoothing (4.6) with smoothing scale ω_{sm}. Let $\varepsilon_{i0\tau}^{[s]} \sim$ EV1 denote the draw-specific base-alternative idiosyncratic shock and normalize the deterministic utility of the base alternative to zero. For $j \geq 1$, define

$$\widetilde{u}_{ij\tau}^{[s]}(\delta, \vartheta) \equiv \frac{\bar{u}_{ij\tau}(\delta, \vartheta) + \nu_{ij\tau}^{[s]} - \varepsilon_{i0\tau}^{[s]}}{\omega_{\text{sm}}}.$$

Then the smoothed choice probabilities are

$$P_{ij\tau}^{[s]}(\delta, \vartheta) = \frac{a_{j,m(i,\tau)} \exp\left(\widetilde{u}_{ij\tau}^{[s]}(\delta, \vartheta)\right)}{1 + \sum_{k=1}^{J} a_{k,m(i,\tau)} \exp\left(\widetilde{u}_{ik\tau}^{[s]}(\delta, \vartheta)\right)}, \qquad P_{i0\tau}^{[s]}(\delta, \vartheta) = \frac{1}{1 + \sum_{k=1}^{J} a_{k,m(i,\tau)} \exp\left(\widetilde{u}_{ik\tau}^{[s]}(\delta, \vartheta)\right)}.$$

Using common random numbers across occasions, define the draw-specific probability of the observed outcome by

$$p^0_{i\tau}(\theta_\delta) \equiv p^0_{i\, j_{i\tau}\, \tau}(\theta_\delta) = \sum_{j=0}^{J} y_{ij\tau}\, p^0_{ij\tau}(\theta_\delta).$$

and hence the simulated singleton and pair probabilities

$$\widehat{p}_{i\tau}(\delta,\vartheta) \equiv \frac{1}{S}\sum_{s=1}^{S} P^{[s]}_{i\tau}(\delta,\vartheta), \qquad \widehat{p}_{i\tau\tau'}(\delta,\vartheta) \equiv \frac{1}{S}\sum_{s=1}^{S} P^{[s]}_{i\tau}(\delta,\vartheta)\, P^{[s]}_{i\tau'}(\delta,\vartheta).$$

Let

$$g_{i\tau\tau'}(\delta,\vartheta) \equiv \log\left(\frac{\widehat{p}_{i\tau\tau'}(\delta,\vartheta)}{\widehat{p}_{i\tau}(\delta,\vartheta)\,\widehat{p}_{i\tau'}(\delta,\vartheta)}\right)$$

denote the pairwise dependence log-ratio. We evaluate the dependence correction on a random subsample of within-household pairs $\mathcal{P}_i \subset \{(\tau,\tau') : \tau < \tau',\ \tau,\tau' \in \mathcal{T}_i\}$. Let $N_{\text{cases}} \equiv \sum_i |\mathcal{T}_i|$ denote the number of observed micro choice occasions and let $|\mathcal{P}| \equiv \sum_i |\mathcal{P}_i|$ denote the total number of sampled pairs. We rescale the pooled pair sum by

$$w \equiv \frac{N_{\text{cases}}}{2|\mathcal{P}|},$$

which keeps the dependence term on an $O(N_{\text{cases}})$ scale even though the number of candidate pairs grows quadratically in $|\mathcal{T}_i|$.

Micro-to-macro expansion weight for the dependence term. The dependence correction is a per-household object: in a full-population panel it would be summed over all consumers whose repeated choices generate the macro shares. Extending the micro–macro pseudo-likelihood logic in Grieco et al. (2025) to panel micro data, we therefore scale the dependence term by an expansion factor that maps the micro panel to the underlying population. Let

$$N^{\text{tot}}_m \equiv \sum_{j=0}^{J} N^{\text{tot}}_{jm}, \qquad O^{\text{pop}} \equiv \sum_{m\in\mathcal{M}} N^{\text{tot}}_m.$$

Under the approximation that the average number of observed purchase occasions per consumer is similar in the micro panel and in the underlying population, a convenient empirical proxy for the expansion factor is

$$\Omega_{\text{raw}} \equiv \frac{O^{\text{pop}}}{N_{\text{cases}}}. \tag{C.3}$$

Because Ω_{raw} can be large in applications and the dependence correction is itself a composite-likelihood object, our baseline implementation uses a conservative expansion weight that leaves the in-sample micro dependence term at unit weight and down-weights only the extrapolated portion:

$$\Omega_{\text{dep}} \equiv 1 + \varsigma\,(\Omega_{\text{raw}} - 1)\,, \qquad \varsigma = 0.01. \tag{C.4}$$

Putting these pieces together, the micro component used in estimation is

$$\mathcal{L}_{\text{micro}}(\delta, \vartheta) = \sum_i \sum_{\tau \in \mathcal{T}_i} \log \widehat{p}_{i\tau}(\delta, \vartheta) + \Omega_{\text{dep}}\, w \sum_i \sum_{(\tau,\tau') \in \mathcal{P}_i} g_{i\tau\tau'}(\delta, \vartheta).$$

C.4 Instruments and the PLM penalty

To address price endogeneity, we impose orthogonality between excluded instruments and the within-alternative component of mean utility. Let z_{jm} denote a Hausman-style instrument (the mean price of the same alternative in other markets in the same week), and let $\widetilde{Z}$ denote a matrix collecting polynomial transforms of z_{jm} after demeaning within alternative and standardization. Let $\widetilde{\delta}_{jm} \equiv \delta_{jm} - \bar{\delta}_j$ denote within-alternative deviations in mean utility, so alternative fixed effects are treated as included controls. Stacking $\widetilde{\delta}$ and $\widetilde{Z}$ over observed (j, m) cells, the sample moments are $\widehat{m}(\delta) \equiv \widetilde{Z}'\widetilde{\delta}$ and the PLM penalty is

$$\chi_{\text{PLM}}(\delta) \equiv \frac{1}{2}\, \widehat{m}(\delta)' W\, \widehat{m}(\delta), \tag{C.5}$$

with W positive definite (we use a first-step weight based on $(\widetilde{Z}'\widetilde{Z})^{-1}$).

C.5 Standard errors for panel-CLEER with AR–persistent errors

This appendix describes inference for the panel-CLEER objective in (5.7). Because the micro component (5.6) is a *composite* (pairwise) likelihood, the information matrix equality generally fails (Varin et al., 2011). Furthermore, both the micro panel and the macro data exhibit within-cluster dependence. We therefore rely on a cluster-robust Godambe/sandwich variance estimator, following recent applications to autoregressive panel probit models (Godambe, 1960; Huber, 1967; Tuzcuoglu, 2023).

Objective and cluster-robust variance. Let $\psi \equiv \left(\delta', \theta'\right)'$ with $\theta \equiv (\lambda_P, \lambda_D)'$, and denote the sample objective as $\Omega(\psi) \equiv \mathcal{L}_{\text{macro}} + \mathcal{L}_{\text{micro}} - \tau_{PLM}\, \chi_{\text{PLM}}$. Let $\widehat{H} \equiv -\nabla^2_{\psi}\Omega(\widehat{\psi})$ be the observed Hessian. The clustered Godambe variance is given by

$$\widehat{\text{Var}}(\widehat{\psi}) = \widehat{H}^{-1}\, \widehat{J}\, \widehat{H}^{-1}, \qquad \widehat{J} \equiv \sum_{c} \widehat{s}_c \widehat{s}_c', \tag{C.6}$$

where $\widehat{s}_c$ is the cluster-level score evaluated at $\widehat{\psi}$.

Cluster score contributions. For the micro component, we cluster by household i to account for overlapping within-household pairs. For the macro and PLM components, we cluster by store rather than by market cell: all store–week contributions belonging to the same store are aggregated before forming the outer product of scores. Let $\mathcal{M}_s \equiv \{m \in \mathcal{M} : \text{store}(m) = s\}$. Then the macro/PLM score contribution for store s is obtained by summing the corresponding market-level macro and PLM score contributions over $m \in \mathcal{M}_s$. The sandwich matrix is therefore

$$\widehat{J} = \sum_{i} \widehat{s}_i^{\text{micro}} \widehat{s}_i^{\text{micro}\prime} + \sum_{s} \widehat{s}_s^{\text{M+P}} \widehat{s}_s^{\text{M+P}\prime}.$$

To reduce the computational burden of inverting the full high-dimensional matrix $\widehat{H}$, one can use the Schur complement to concentrate out the mean utilities δ and directly obtain the variance block for the low-dimensional parameters θ (Conlon and Gortmaker, 2020; Grieco et al., 2025). Finally, given the estimated variance for $\widehat{\delta}$, the variance for the linear parameters $\widehat{\beta}$ (including the price coefficient $\widehat{\alpha}$) is obtained via the standard delta method.

D Welfare analysis in the AR-persistent specification

This appendix describes how we compute changes in consumer surplus for counterfactuals in the AR-persistent specification in (4.2). The key observation is that, conditional on the simulated latent Cardell components, each purchase occasion is a standard multinomial logit with scale parameter Λ; therefore the usual inclusive value formula applies inside each simulation draw.

D.1 Setup and notation

Let $r \in \{0, 1\}$ index two environments (e.g., $r = 0$ baseline and $r = 1$ counterfactual). At purchase occasion t, consumer i faces availability indicators $\{a^r_{jt}\}_{j=0}^{J}$ and systematic utilities $\{V^r_{ijt}\}_{j=0}^{J}$ (which

may differ across r because prices, availability, or product characteristics differ). Utility is

$$u_{ijt}^r = V_{ijt}^r + \epsilon_{ijt}, \qquad j \in \{0, 1, \ldots, J\},\ t \in \{1, \ldots, T\}. \tag{D.1}$$

In the AR-persistent model, the unobserved component is (restating (4.2))

$$\epsilon_{ijt} = \underbrace{\nu_{ij}}_{\text{persistent}} + \underbrace{\lambda_P \sum_{\ell=0}^{L} \lambda_D^{\ell} \nu_{ij,t-\ell}}_{\text{truncated AR component}} + \underbrace{\Lambda\, \varepsilon_{ijt}}_{\text{i.i.d. EV1 shock}}, \qquad \Lambda \equiv \lambda_P \lambda_D^L, \tag{D.2}$$

where $\varepsilon_{ijt} \overset{iid}{\sim} \text{EV1}$ and $\nu_{ij} \overset{iid}{\sim} C(\lambda_P)$, $\nu_{ijt} \overset{iid}{\sim} C(\lambda_D)$, mutually independent across all indices.

Define the *structured* latent component (the part we simulate)

$$\eta_{ijt} \;\equiv\; \nu_{ij} + \lambda_P \sum_{\ell=0}^{L} \lambda_D^{\ell} \nu_{ij,t-\ell}, \qquad \text{so that} \qquad \epsilon_{ijt} = \eta_{ijt} + \Lambda\, \varepsilon_{ijt}. \tag{D.3}$$

D.2 Change in consumer surplus conditional on latent shocks

Fix (i, t) and condition on a realization $\eta_{it} \equiv (\eta_{i0t}, \ldots, \eta_{iJt})$. The only remaining randomness in $\{u_{ijt}^r\}_j$ is $\{\Lambda \varepsilon_{ijt}\}_j$, which is i.i.d. EV1 with scale Λ. Therefore the conditional expected maximum utility is given by:

$$\mathbb{E}\left[\max_{j:\, a_{jt}^r = 1} u_{ijt}^r \,\middle|\, \eta_{it} \right] = \Lambda \log\left(\sum_{j:\, a_{jt}^r = 1} \exp\left(\frac{V_{ijt}^r + \eta_{ijt}}{\Lambda} \right) \right) + \Lambda\, \gamma, \tag{D.4}$$

where γ is the Euler constant.

Define the (conditional) inclusive value

$$IV_{it}^r(\eta_{it}) \;\equiv\; \log\left(\sum_{j:\, a_{jt}^r = 1} \exp\left(\frac{V_{ijt}^r + \eta_{ijt}}{\Lambda} \right) \right). \tag{D.5}$$

Then (D.4) becomes $\mathbb{E}[\max u_{ijt}^r \mid \eta_{it}] = \Lambda I V_{it}^r(\eta_{it}) + \Lambda\gamma$. Let $\alpha_i > 0$ denote the marginal disutility of price. Hence, conditional on η_{it}, the per-occasion change in consumer surplus from moving from

environment 0 to 1 is

$$\Delta CS_{it}(\eta_{it}) \equiv \frac{1}{\alpha_i}\left(\mathbb{E}\left[\max_{j:\,a^1_{jt}=1} u^1_{ijt}\,\middle|\,\eta_{it}\right] - \mathbb{E}\left[\max_{j:\,a^0_{jt}=1} u^0_{ijt}\,\middle|\,\eta_{it}\right]\right)$$
$$= \frac{\Lambda}{\alpha_i}\left(IV^1_{it}(\eta_{it}) - IV^0_{it}(\eta_{it})\right). \tag{D.6}$$

D.3 Simulation over the persistent and AR components

Because η_{ijt} in (D.3) is random and serially correlated over t, we compute ex-ante expected surplus changes by simulation over the Cardell components, using the same integration over logit choice probabilities as in the likelihood.

For each consumer i, draw $s = 1, \ldots, S$ realizations of the primitive shocks $\{\nu^{[s]}_{ij}, \nu^{[s]}_{ijt}\}$ and construct

$$\eta^{[s]}_{ijt} = \nu^{[s]}_{ij} + \lambda_P \sum_{\ell=0}^{L} \lambda^{\ell}_D \nu^{[s]}_{ij,t-\ell}, \qquad t = 1, \ldots, T,\ j = 0, 1, \ldots, J.$$

For each draw and each environment $r \in \{0, 1\}$, compute the inclusive value

$$IV^{r,[s]}_{it} = \log\left(\sum_{j:\,a^r_{jt}=1} \exp\left(\frac{V^r_{ijt} + \eta^{[s]}_{ijt}}{\Lambda}\right)\right).$$

Then the simulated per-occasion welfare change is

$$\widehat{\Delta CS}_{it} \;=\; \frac{1}{S}\sum_{s=1}^{S} \frac{\Lambda}{\alpha_i}\left(IV^{1,[s]}_{it} - IV^{0,[s]}_{it}\right). \tag{D.7}$$

In practice, we use the same underlying random draws to construct $\eta^{[s]}_{ijt}$ in both environments $r = 0$ and $r = 1$. This delivers a low-variance estimator of the difference in (D.7) and corresponds to holding fixed the consumer's latent tastes and idiosyncratic history when evaluating the counterfactual.

E OTC cough remedy summary statistics

Table E.1: Summary Statistics for NielsenIQ Scanner Data Cough Medicine Sample

	N	Mean	SD	P25	P50	P75	Min	Max
Share	278159	0.103	0.098	0.038	0.071	0.132	0.004	1.000
Price	278159	10.930	4.781	7.490	9.990	12.990	0.010	51.990
Price (relative to base)	278159	3.442	4.867	0.000	2.500	5.750	-13.280	47.470
Hausman IV	278159	10.771	3.866	7.382	10.101	13.947	0.393	25.348
Hausman IV (rel)	278159	3.437	3.892	0.000	2.848	6.644	-7.524	19.595

Notes: This table reports summary statistics for the market-level scanner data sample (NielsenIQ RMS) for the years 2011–2013. An observation is a store–week–alternative. The choice set consists of the top ten brand–formulation pairs and a composite "All Other" alternative that aggregates remaining products and private labels. Share represents the within-category market share, as the model conditions on category purchase. Prices are standardized unit prices. The Hausman IV is the quantity-weighted mean unit price of the same alternative in all other Nielsen DMAs during the same week. Variables denoted "relative to base" are differenced from the composite alternative.

Table E.2: Summary Statistics for Household Survey Data Cough Medicine Sample

	Mean	SD	P25	P50	P75	Min	Max
Price	10.935	4.780	7.490	9.990	12.990	0.010	51.990
Price (relative to base)	3.451	4.865	0.000	2.509	5.754	-13.280	47.470
Share	0.101	0.096	0.038	0.071	0.129	0.004	1.000
Product Characteristics × Demographics (Mean Utility Shifters)							
Liquid × older	0.345	0.461	0.000	0.000	1.000	0.000	1.000
Liquid× younger	0.017	0.127	0.000	0.000	0.000	0.000	1.000
Lozenge × older	0.010	0.098	0.000	0.000	0.000	0.000	1.000
Lozenge × younger	0.000	0.019	0.000	0.000	0.000	0.000	1.000
Tablet × older	0.006	0.050	0.000	0.000	0.000	0.000	0.400
Tablet × younger	0.000	0.010	0.000	0.000	0.000	0.000	0.400
Household Characteristics (Price Sensitivity Shifters)							
Log income	10.865	0.658	10.532	10.915	11.350	7.824	11.631
HH head age	56.968	12.790	48.000	57.000	66.000	19.000	111.000
Household size	2.538	1.293	2.000	2.000	3.000	1.000	9.000
Has young child	0.447	0.328	0.447	0.447	0.447	0.000	1.000
2+ working adults	0.066	0.164	0.000	0.066	0.066	0.000	1.000
Households				14,611			
Trips				27,228			
Avg alts per trip				9.891			
N				269,310			

Notes: This table reports summary statistics for the household-level sample (NielsenIQ Homescan) for the years 2011–2013. The sample includes 27,228 category purchase trips made by 14,611 distinct households. The choice set for each trip is defined by the alternatives available in the linked store–week, utilizing a rolling carried indicator. Mean utility shifters consist of product formulations interacted with local market demographics. Household characteristics serve as price sensitivity shifters in the demand model. While we use standardized measures in the model, they are reported here in their unstandardized form. Summary statistics for prices and shares are computed across all inside alternatives presented in the consumers' choice sets.