Dynamic Discrete Choice Copula Models

Andrew Shephard*

February 6, 2026

Preliminary and incomplete. Please do not circulate or cite without permission.

Abstract

A copula-based specification is introduced to model serial dependence in utility shocks while preserving logit tractability in dynamic discrete choice models. With Type-I extreme value marginals and a Bernstein-polynomial copula, persistence is summarized by an endogenous finite-state latent rank Markov process: conditional choice probabilities and inclusive values are closed-form finite mixtures of logit expressions, and the likelihood integrates out the latent state by standard hidden Markov model filtering.

Keywords: dynamic discrete choice; serial dependence; copulas; Bernstein polynomials; hidden Markov models; conditional choice probabilities.

1 Introduction

Economic behaviour is often persistent: choices, prices, and allocations display inertia because incentives and constraints evolve slowly, information accumulates, and unobserved conditions remain correlated over time. In this paper we focus on persistence in dynamic discrete choice models, where agents make sequential choices from a finite set of alternatives while taking into account future consequences. Such models are widely used in empirical economics to study intertemporal decision-making in contexts such as labour supply and occupational choice (Eckstein and Wolpin, 1989; Keane and Wolpin, 1997), dynamic demand and product choice under uncertainty (Erdem and Keane, 1996), and firm investment and industry dynamics (Rust, 1987; Ericson and Pakes, 1995).

In the canonical dynamic discrete choice framework,[1] every period agents observe an information set consisting of an observed state and a vector of unobserved utility shocks. They then choose an alternative to maximise expected discounted utility, taking into account how current choices affect future states and payoffs. The unobserved shocks capture idiosyncratic tastes, costs, or other factors that vary across alternatives and over time. A common assumption in the literature is that these shocks are independent and identically distributed (i.i.d.) over time with Type-I extreme value marginals, which yields the familiar logit choice

*KU Leuven. E-mail: andrew.shephard@kuleuven.be.

[1]Recent surveys are provided in Aguirregabiria and Mira (2010), Keane et al. (2011) and Arcidiacono and Ellickson (2011).

probabilities and closed-form expressions for inclusive values (McFadden, 1978). As shown by Rust (1994), these assumptions greatly simplify the solution and estimation of dynamic discrete choice models.

While it has long been recognised that allowing for serial dependence in unobserved components is important for more realistic modelling of dynamic choice behaviour, the reliance on i.i.d. shocks remains widespread and often motivated by computational tractability rather than necessarily empirical plausibility (e.g., Rust and Phelan, 1997). In this paper we revisit the issue of serial dependence in dynamic discrete choice models, and propose a novel copula-based approach (Sklar, 1959) that retains much of the computational convenience of the classic logit framework while allowing for flexible persistence in the unobserved components. In contrast to approaches that accommodate serial dependence through simulation or numerical approximation of the dynamic program and the likelihood, our method is exact, delivering key objects in closed form and requiring neither simulation nor approximation of the dynamic programming problem.

Persistence in observed choices can arise through distinct economic mechanisms. One channel is *state dependence*: past choices change future payoffs through the observed state (e.g., experience accumulation, switching costs, or learning). A second channel is persistence in *unobserved* components: tastes, costs, or wage shocks that are serially correlated over time. These channels have different counterfactual implications and are often hard to disentangle empirically (see, e.g., Heckman, 1981; Hyslop, 1999). This paper focuses on the second channel, and provides a tractable way to introduce serial correlation in utility shocks while preserving the familiar logit building blocks.

The challenges of introducing such serial dependence in dynamic discrete choice models are well known. In their survey, Aguirregabiria and Mira (2010) emphasise two related sources of complexity. First, once unobservables are persistent, the observed state is generally not a sufficient statistic for the agent's information: likelihood contributions need not factor into conditional choice probabilities given the observables alone, and evaluation requires integrating over latent state histories compatible with the observed data. Second, when the relevant persistent components are continuous, the dynamic programming problem and the likelihood typically require approximation. For example, through interpolation or polynomial methods for the value function and through simulation-based or quadrature-based integration over latent paths.

A number of papers have confronted these challenges. Norets (2009) develops a Bayesian framework for dynamic discrete choice models where serially correlated unobservables are present, demonstrating that Markov chain Monte Carlo (MCMC) methods can handle the resulting high-dimensional integration in the likelihood function. Blevins (2016) proposes a simulation-based maximum likelihood estimator using sequential Monte Carlo (particle filtering) to approximate the likelihood function. Both approaches are flexible and general, but they rely on simulation methods. Deterministic integration methods, such as Gaussian quadrature, have also been proposed to approximate the integrals over unobserved states (Stinebrickner, 2000). More recently, Reich (2018) develops a method that decomposes the high-dimensional integral into a sequence of lower-dimensional integrals. Non-full solution methods, such as conditional choice probability (CCP) approaches, have also been adapted to settings with serially correlated unobservables (Arcidiacono and Miller, 2011).

This paper proposes a complementary approach that uses a parametric copula model to introduce serial dependence.[2] While there are many possible copula choices that can be used to model serial dependence (see, for example, Nelsen, 2006 and Joe, 2014), we argue that the Bernstein-polynomial copula (Sancetta and Satchell, 2004) when paired with Type-I extreme value marginals, is particularly well suited to our setting. Our main technical result is that this pairing yields *exact* closed-form within-period objects conditional on a low-dimensional sufficient statistic (latent rank state). In particular, conditional choice probabilities and inclusive values become finite mixtures of standard logit expressions (Proposition 1). The copula construction implies that the latent rank state evolves Markovianly over time with an explicit transition matrix determined by the copula weights. As a result, likelihood evaluation reduces to standard forward filtering in a finite hidden Markov model (HMM). Here, the latent state does not represent an ad hoc "regime", but rather it is an endogenous sufficient statistic for the conditional shock distribution.

The latent state and HMM structure imply a close relationship to "hidden Rust" models. In Connault (2016), persistence is introduced by augmenting the agent's state with an unobserved, discrete Markov component that shifts flow utility, while maintaining i.i.d. Type-I extreme value shocks so that conditional choice probabilities are standard logit given the full (partially unobserved) state. Our approach instead places persistence in the utility shocks themselves, with the HMM representation emerging endogenously as a sufficient statistic for the conditional shock distribution. Our latent state has a clear statistical interpretation as a rank index of the current shock within its marginal distribution. Persistence means that if today's shock was in the upper tail, tomorrow's shock is also more likely to be in the upper tail. In both cases likelihood evaluation reduces to forward filtering over a finite latent state, but the objects being filtered differ: hidden Rust filters an unrestricted latent "utility regime" with label permutations a priori observationally equivalent, whereas our latent ranks come with a natural ordering and a copula-implied transition law.

This structure also clarifies how identification results from the existing literature apply in our setting. Hu and Shum (2012) establish nonparametric identification in dynamic models with a serially correlated latent state, showing that the joint Markov law of motion is identified from short panels under a Markov structure, rank conditions, and an operator invertibility/completeness condition. Connault (2016) studies dynamic algebraic models (including parametric dynamic discrete choice models with unobserved state variables) and proves two generic identification results: (1) the identification structure is constant almost everywhere on the parameter space (outside an exceptional, measure-zero set), and (2) the identification structure stabilizes at a finite horizon, so identification from the infinite-horizon distribution implies identification from a finite collection of finite-horizon marginals.[3]

[2]A copula is a function that *couples* marginal distributions to describe their joint dependence structure independently of the individual variables' behaviours. For an overview of copula methods and applications in economics, see Patton (2012) and Fan and Patton (2014).

[3]A related literature considers identification in the context of time-invariant unobserved heterogeneity ("types"), leading to finite mixture models (Kasahara and Shimotsu, 2009).

Our model inherits this structure. Viewed as a latent-state choice model, the data identify (i) state-dependent conditional choice probabilities and (ii) the latent transition law, up to relabelling of the latent states. Given Type-I extreme value marginals and the copula restriction, these reduced-form objects then map back to economically interpretable primitives. Because the latent states correspond to rank regions, they come with a natural ordering that is convenient for labelling.

The Bernstein degree $m \geq 1$ serves as a transparent tuning parameter controlling the flexibility of serial dependence: when $m = 1$ shocks are i.i.d. and we recover the classic logit model, while larger m yields increasingly flexible serial dependence patterns (and, as m grows, Bernstein copulas can approximate arbitrary copulas under mild regularity conditions). We also propose low-dimensional parameterizations of the copula weights that map directly into interpretable dependence measures, such as Spearman's rank correlation. Under injectivity of the mapping from copula parameters to the latent transition matrix, the structural parameters are identified from the observed data distribution via the identified reduced-form objects.

To demonstrate the approach, we first present a simulation study of a simple occupational choice model, where we show how serial dependence in unobservables can generate persistent choice behaviour beyond what is explained by observed states alone. We use this example to illustrate identification and to show that ignoring serial dependence can lead to biased structural estimates. We then revisit the seminal bus engine replacement model of Rust (1987), allowing for serial dependence in unobservables.[4] In this application we find substantial persistence in unobserved maintenance costs, with the estimated dependence implying Spearman rank correlations on the order of 0.6–0.8 for shocks one period apart (depending on the copula degree and the discount factor). We also compare the results to the standard i.i.d. logit model and to alternative methods that accommodate serial dependence through simulation-based techniques.

The remainder of the paper is organised as follows. In Section 2 we set up the dynamic discrete choice model with serially dependent shocks. Section 3 introduces the copula framework and the Bernstein copula specification, derives closed-form expressions for conditional choice probabilities and inclusive values, and describes likelihood evaluation via filtering over the latent state. We then illustrate our approach with two complementary examples: in Section 4 we present an occupational choice simulation study, while Section 5 revisits the seminal bus engine replacement application of Rust (1987) allowing for serial dependence in unobservables. Finally, Section 6 concludes. Proofs and additional results are collected in the Appendix.

[4] While the Rust model is characterized by its relative simplicity and tractability, its structure captures the main features of more complicated models. It has become a standard benchmark for demonstrating new methods in the dynamic discrete choice literature. See, for example, Aguirregabiria and Mira (2002), Bajari et al. (2007), Norets (2009), Arcidiacono and Miller (2011), Su and Judd (2012), Larsen et al. (2012), Norets and Tang (2013), Chiong et al. (2016), Blevins (2016), Reich (2018), Kristensen et al. (2021), and Norets and Shimizu (2024).

2 Dynamic discrete choice with serially dependent shocks

Consider I alternatives and a dynamic discrete choice problem over periods $t = 1, \ldots, T$, where the horizon T may be finite or infinite, and let $\beta \in (0, 1)$ denote the discount factor. Let x_t denote the observed state and let $y_t \in \{1, \ldots, I\}$ denote the choice. Each alternative j has an additive, alternative-specific utility shock ε_{jt} in period t, collected in the vector $\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{It})$.

Assumption AS (Additive separability)**.** *Per-period utility is additively separable and given by*

$$U_t(x_t, y_t, \varepsilon_t) = u(x_t, y_t) + \varepsilon_{y_t t}.$$

The state x_t and shock vector ε_t are observed by the agent at time t, while the econometrician observes x_t and the choice y_t.

The key feature of our setting relative to Rust (1987) and related frameworks, is that shocks are serially dependent. This means that the distribution of future shocks depends on current realizations. Writing $V_t(x_t, \varepsilon_t)$ for the value function, the decision rule can be written as

$$y_t \in \arg\max_{j \in \{1,\ldots,I\}} \Big\{ u(x_t, j) + \varepsilon_{jt} + \beta\, \mathbb{E}\big[V_{t+1}(x_{t+1}, \varepsilon_{t+1}) \mid x_t, y_t = j, \varepsilon_t \big] \Big\}. \tag{1}$$

Because ε_{t+1} is serially dependent with ε_t, the continuation term in (1) is, in general, a function of the entire vector ε_t. For the evolution of the observed state, we will maintain a conditional-independence restriction that allows dependence on a low-dimensional latent dependence state (constructed below), but rules out any direct dependence on the full contemporaneous shock vector once we condition on that state (Assumption CI). In the $T < \infty$ case, the value function satisfies the terminal condition $V_{T+1} \equiv 0$.[5]

We work under the following standard assumption on the marginals and within-period independence of the shocks.

Assumption CL (Marginals and within-period independence)**.** *For each t, $(\varepsilon_{1t}, \ldots, \varepsilon_{It})$ are independent and identically distributed with Type-I extreme value (T1EV) marginal CDF*

$$F(\varepsilon) = \exp\{-\exp(-\varepsilon)\}.$$

In a purely static model, Assumptions AS and CL imply the familiar logit formula for CCPs (see, McFadden, 1978). In a dynamic model with serial dependence, however, T1EV marginals by themselves do not guarantee that CCPs conditional on the agent's observed information are logit. The reason is that serial dependence can make the conditional law of the current shock vector, given observables and/or past choices (and hence past shocks), depart from i.i.d. T1EV across alternatives; in dynamic settings this can happen even when shocks are independent across alternatives within a period. As a result, away from special

[5] If $T = \infty$ and the environment is stationary, one can write the value function without time subscripts as $V(x, \varepsilon)$ and interpret (1) as the optimal policy rule induced by the stationary Bellman equation.

cases, one should not infer logit CCPs from marginal T1EV assumptions alone.

Our goal is to introduce serial dependence in a way that yields a low-dimensional latent state whose evolution is Markovian, and conditional on that state, the resulting CCPs and inclusive values admit tractable finite-sum closed forms.

3 Encoding serial dependence with copulas

A copula model (Sklar, 1959) constructs multivariate distributions by combining marginal distributions with a dependence structure. Let $U_{jt} = F(\varepsilon_{jt}) \in (0, 1)$. By construction, each U_{jt} is marginally uniform. We assume alternative-wise independence, while serial dependence for a given alternative j is encoded by a bivariate copula C linking $(U_{j,t-1}, U_{jt})$.

Assumption SC (Serial copula, alternative-wise)**.** *For each j, $(U_{j,t-1}, U_{jt})$ has copula C with density c on $(0, 1)^2$. Across alternatives, the pairs are independent*

$$(U_{1,t-1}, U_{1t}), \ldots, (U_{J,t-1}, U_{Jt}) \text{ are independent.}$$

Assumption SC isolates persistence to "same-alternative over time" dependence. It is a natural baseline for dynamic discrete choice models where the unobserved attractiveness of each alternative is persistent, but there is no common factor linking alternatives across time. As any adjacent-pair copula can be extended (via the copula Markov product) to a coherent family of multi-period distributions (Darsow et al., 1992), the modeling exercise is not existence, but tractability. Namely, how to choose C so that the resulting dynamic discrete choice model admits closed-form CCPs and inclusive values conditional on a low-dimensional latent state.

3.1 A Bernstein copula for serial dependence

We now impose a copula specification that (i) allows flexible serial dependence, while (ii) retaining a finite-mixture structure that will deliver closed-form CCPs, inclusive values, and latent-state transitions. For this purpose, we consider the *Bernstein* copula (Sancetta and Satchell, 2004).[6]

Assumption BC (Bernstein copula for serial dependence)**.** *Fix an integer degree $m \geq 1$ and let $B_{r,m}(u) \equiv \mathbb{P}\{\text{Beta}(r, m-r+1) \leq u\} = \sum_{s=r}^{m} \binom{m}{s} u^s (1-u)^{m-s}$ denote the Beta CDF for $r = 1, \ldots, m$. For each alternative j, the serial copula for $(U_{j,t-1}, U_{jt})$ is the Bernstein mixture*

$$C(u, v) = \sum_{r=1}^{m} \sum_{s=1}^{m} w_{rs} \, B_{r,m}(u) \, B_{s,m}(v), \tag{2}$$

[6] Sancetta and Satchell (2004) propose a Bernstein copula directly in the Bernstein basis (Lorentz, 1986). This Beta-kernel representation of the Bernstein copula is standard in the nonparametric "empirical beta copula" literature. See, for example, Segers et al. (2017).

where $W \equiv (w_{rs})$ *satisfies*

$$w_{rs} \geq 0, \qquad \sum_{s=1}^{m} w_{rs} = \frac{1}{m}\ (\forall r), \qquad \sum_{r=1}^{m} w_{rs} = \frac{1}{m}\ (\forall s). \tag{3}$$

We will now argue why the Bernstein copula is a particularly natural choice for our setting. Some of these points are standard in the copula literature; others are more specific to our dynamic discrete choice application.

Flexibility. The Bernstein copula is a rich nonparametric class that can approximate a wide range of dependence patterns by increasing the degree m. Sancetta and Satchell (2004) show that Bernstein copulas are dense in the space of continuous copulas on $[0,1]^2$ under the uniform norm. Thus, for any continuous copula C^* and any $\epsilon > 0$, there exists a degree m and weights $\{w_{rs}\}$ satisfying (3) such that $\sup_{(u,v)\in[0,1]^2} |C(u,v) - C^*(u,v)| < \epsilon$. Furthermore, a degree-$m$ Bernstein copula is exactly nested in any higher degree $m' > m$ through standard Bernstein degree elevation identities.

Feasibility constraints. The constraints in (3) are the central feasibility advantage. The validity constraints reduce to nonnegativity and simple marginal (row/column-sum) constraints on $\{w_{rs}\}$. Equivalently, mW is a doubly-stochastic matrix, so the feasible set is a (scaled) Birkhoff polytope. This contrasts with many alternative polynomial copula classes where validity constraints are considerably more complex to impose (see, Nelsen, 2006).[7]

Latent-rank representation. Many copula models do not admit a natural latent-index representation. The conditional law $U_t \mid U_{t-1} = u$ typically depend on the continuous value u such that the continuation value in (1) remains a function of U_{t-1} (a continuum), not a finite index. Fortunately, Bernstein copulas admit a natural latent-index representation based on order statistics. This representation is key to tractability in our dynamic discrete choice setting, as it yields a finite-state Markov chain that captures all serial dependence. In particular, for each alternative j there exist latent indices $(K_{j,t-1}, K_{jt}) \in \{1,\ldots,m\}^2$ such that

$$\mathbb{P}\{K_{j,t-1} = r, K_{jt} = s\} = w_{rs},$$

and conditional on $(K_{j,t-1}, K_{jt}) = (r,s)$, the pair $(U_{j,t-1}, U_{jt})$ is independent with

$$U_{j,t-1} \mid K_{j,t-1} = r \sim \text{Beta}(r, m-r+1), \qquad U_{jt} \mid K_{jt} = s \sim \text{Beta}(s, m-s+1).$$

[7] A common polynomial approach is to expand the copula *density* in an orthonormal basis on $[0,1]$. For example, choose functions $\{\phi_r\}_{r=1}^m$ with $\int_0^1 \phi_r(u)\,\mathrm{d}u = 0$ and consider

$$c(u,v) = 1 + \sum_{r=1}^{m} \theta_r\, \phi_r(u)\, \phi_r(v),$$

which preserves uniform marginals by construction. Validity requires the semi-infinite constraint $c(u,v) \geq 0$ for all $(u,v) \in [0,1]^2$. For e.g., Legendre or Chebyshev expansions this does not reduce to simple box constraints on coefficients. One straightforward approach is to enforce the constraint on a grid of points, which turns the semi-infinite constraint into a finite set of linear inequalities. However, this only guarantees non-negativity at the grid points, not everywhere.

Moreover, the Bernstein marginal constraints imply that, for the bivariate pair $(K_{j,t-1}, K_{jt})$ associated with the copula for $(U_{j,t-1}, U_{jt})$, both marginals are uniform on $\{1, \ldots, m\}$. In the induced Markov interpretation, this means the uniform distribution is stationary. From Bayes' rule, we have

$$\mathbb{P}\{K_{jt} = s \mid K_{j,t-1} = r\} = \pi_{rs} \equiv m\, w_{rs}, \qquad r, s \in \{1, \ldots, m\}.$$

This latent-rank structure clarifies how the Bernstein copula in (2) encodes serial dependence. Recall that if $\widetilde{U}_1, \ldots, \widetilde{U}_m \overset{\text{iid}}{\sim} \text{Unif}(0,1)$ and $\widetilde{U}_{(r)}$ denotes the r^{th} order statistic, then $\widetilde{U}_{(r)} \sim \text{Beta}(r, m-r+1)$. Thus, conditional on $K_{j,t-1} = r$, the draw $U_{j,t-1}$ can be interpreted as a randomly selected r^{th} order statistic from an (implicit) sample of size m; likewise for U_{jt} given $K_{jt} = s$. As the rank r increases, more mass is shifted to higher values of $U_{j,t-1}$ (similarly for rank s and U_{jt}). Similarity in rank across periods ($r \approx s$) therefore induces positive dependence in $(U_{j,t-1}, U_{jt})$, while diffuse ranks imply mobility. The transitions between ranks over time, governed by the matrix $\Pi \equiv (\pi_{rs})$, determine the serial dependence structure.[8]

Having established the existence of a latent-index dependence state, we now provide the conditional-independence restriction for the evolution of the observed state.

Assumption CI (Observed-state transition given K_t)**.** *The transition of the observed state is conditionally independent of the contemporaneous utility shocks once we condition on (x_t, y_t, K_t). That is, for any measurable set $\mathcal{X}$,*

$$\mathbb{P}\{x_{t+1} \in \mathcal{X} \mid x_t, y_t, K_t, \varepsilon_t\} = \mathbb{P}\{x_{t+1} \in \mathcal{X} \mid x_t, y_t, K_t\}.$$

If x_t is continuous, interpret these probabilities as statements about conditional densities.

Closed form within-period objects. While other copulas also admit latent-index representations (e.g., finite mixtures of Gaussian copulas), the Bernstein copula is particularly convenient in terms of implementation as it is possible to derive tractable closed-form expressions for CCPs and inclusive values, both of which are required for a Rust-style dynamic discrete choice model. The derivation of these is the focus of the next section.

3.2 Closed form solutions for models with serial dependence

This section develops the main ingredients needed to embed serial dependence in a dynamic discrete choice model while preserving tractability. The key steps are: (i) conditioning on K_t to isolate a state that restores within-period tractability; (ii) a change of variables that turns the Beta kernels from the Bernstein copula into finite exponential mixtures, and (iii) a simple identity for scaled exponentials that turns those mixtures into finite-sum CCP and inclusive-value formulas.

[8] This mixture representation also makes it transparent why larger m permits stronger dependence. For fixed latent ranks, the Beta draws become increasingly concentrated as m grows: $U \mid K = k$ has mean $k/(m+1)$ and variance $k(m-k+1)/\{(m+1)^2(m+2)\}$. Hence, increasing m makes $U_{j,t-1}$ and U_{jt} track the same latent rank more tightly.

3.2.1 Preliminaries

Our closed-form solutions follow from the following lemmata.

Lemma 1 (Joint density of Z_t given K_t)**.** *Define $Z_{jt} \equiv -\log U_{jt}$. Under Assumptions CL, SC, and BC, and conditional on $K_t = (K_{1t}, \dots, K_{It})$, the transformed shocks $Z_t \equiv (Z_{1t}, \dots, Z_{It})$ have joint density*

$$f_{Z_t|K_t}(z_1, \dots, z_I) = \sum_{q_1=0}^{m-K_{1t}} \cdots \sum_{q_I=0}^{m-K_{It}} \left[\prod_{i=1}^{I} \kappa_{K_{it}, q_i} \right] \left(\prod_{i=1}^{I} \lambda_{K_{it}, q_i} e^{-\lambda_{K_{it}, q_i} z_i} \right), \tag{4}$$

where the rates $\{\lambda_{sq}\}$ and coefficients $\{\kappa_{sq}\}$ are given by

$$\lambda_{sq} \equiv s + q, \qquad \kappa_{sq} = (-1)^q \binom{m}{s+q} \binom{s+q-1}{q}, \tag{5}$$

with the coefficients satisfying $\sum_{q=0}^{m-s} \kappa_{sq} = 1$.

The proof of Lemma 1 is provided in the Appendix, which also describes a simple recursive procedure to calculate the coefficients. We use this lemma in conjunction with the following standard result for scaled exponentials.

Lemma 2 (Rate shifted logit)**.** *Let $Z_1, \dots, Z_I$ be independent exponentials with rates $\lambda_1, \dots, \lambda_I$ and let $a_1, \dots, a_I$ be deterministic indices. Define the scaled minimum $T \equiv \min_{\ell \in \{1,\dots,I\}} \{Z_\ell e^{-a_\ell}\}$. Then*

$$\mathbb{P}\Big\{ \arg\min_{\ell \in \{1,\dots,I\}} Z_\ell e^{-a_\ell} = j \Big\} = \frac{\lambda_j e^{a_j}}{\sum_{\ell=1}^{I} \lambda_\ell e^{a_\ell}}, \qquad \mathbb{E}\big[-\log T \big] = \gamma + \log \Big(\sum_{\ell=1}^{I} \lambda_\ell e^{a_\ell} \Big), \tag{6}$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant.

3.2.2 Individual optimisation with serially dependent shocks

We now derive closed-form expressions for optimal individual behaviour conditional on the low-dimensional latent state K_t. Define the *choice-specific* value indices

$$\tilde{v}_{jt} \equiv u(x_t, j) + \beta \mathbb{E}\big[V_{t+1}(x_{t+1}, \varepsilon_{t+1}) \mid x_t, y_t = j, K_t \big].$$

By construction, $\tilde{v}_{jt}$ is measurable with respect to (x_t, K_t). Conditioning on K_t ensures that the continuation value does not depend upon the full shock vector ε_t.

Rather than working directly with the original maximisation, define $Z_{jt} \equiv \exp(-\varepsilon_{jt})$ and restate the choice event as an exponential race (Luce and Suppes, 1965). That is, the chosen alternative at time t can be written as

$$y_t \in \arg\min_{\ell \in \{1,\dots,I\}} \{ Z_{\ell t} e^{-\tilde{v}_{\ell t}} \}.$$

Written in this way, the conditional choice probabilities can then be computed by integrating the choice event against the joint density $f_{Z_t|K_t}$. That is

$$\mathbb{P}\{y_t = j \mid K_t, \tilde{\boldsymbol{v}}_t\} = \int_{\mathbb{R}_+^I} \mathbb{1}\{j \in \arg\min_\ell (z_\ell e^{-\tilde{v}_{\ell t}})\}\, f_{Z_t|K_t}(z_1, \ldots, z_I)\, \mathrm{d}z_1 \cdots \mathrm{d}z_I.$$

Lemma 1 provides a finite expansion for $f_{Z_t|K_t}$, so we may exchange the finite sums with the integral and compute the result term-by-term. For each $(q_1, \ldots, q_I)$ term, the coordinates are independent exponentials with rates λ_{K_{it},q_i}. Lemma 2 then yields the logit-like expression for the CCPs as well as the corresponding inclusive-value formula.[9] Averaging over $(q_1, \ldots, q_I)$ with weights $\prod_{i=1}^I \kappa_{K_{it},q_i}$ immediately gives the within-period closed form objects below.

Proposition 1 (Finite-mixture CCPs and inclusive value under the Bernstein copula). *Fix degree $m \geq 1$ and let $K_t \equiv (K_{1t}, \ldots, K_{It}) \in \{1, \ldots, m\}^I$ denote the vector of latent indices at time t. Let $\tilde{\boldsymbol{v}}_t \equiv (\tilde{v}_{1t}, \ldots, \tilde{v}_{It})$ denote the vector of choice-specific value indices. Then, conditional on $(K_t, \tilde{\boldsymbol{v}}_t)$, the period-$t$ CCP is the finite mixture*

$$\mathbb{P}\{y_t = j \mid K_t, \tilde{\boldsymbol{v}}_t\} = \sum_{q_1=0}^{m-K_{1t}} \cdots \sum_{q_I=0}^{m-K_{It}} \left[\prod_{i=1}^I \kappa_{K_{it},q_i}\right] \frac{\lambda_{K_{jt},q_j} \exp(\tilde{v}_{jt})}{\sum_{\ell=1}^I \lambda_{K_{\ell t},q_\ell} \exp(\tilde{v}_{\ell t})}, \tag{7}$$

where $\{\lambda_{sq}\}$ and $\{\kappa_{sq}\}$ are defined in (5). Moreover, define the inclusive value

$$G(K_t, \tilde{\boldsymbol{v}}_t) \equiv \mathbb{E}\Big[\max_{1\leq \ell \leq I} \{\tilde{v}_{\ell t} + \varepsilon_{\ell t}\} \,\Big|\, K_t, \tilde{\boldsymbol{v}}_t\Big].$$

Let $\gamma \approx 0.5772$ denote the Euler–Mascheroni constant. Then

$$G(K_t, \tilde{\boldsymbol{v}}_t) = \gamma + \sum_{q_1=0}^{m-K_{1t}} \cdots \sum_{q_I=0}^{m-K_{It}} \left[\prod_{i=1}^I \kappa_{K_{it},q_i}\right] \log\left(\sum_{\ell=1}^I \lambda_{K_{\ell t},q_\ell} \exp(\tilde{v}_{\ell t})\right). \tag{8}$$

Proposition 1 gives closed-form within-period objects *conditional* on the latent indices K_t. These conditional CCPs and inclusive values are the basic building blocks for (i) solving the dynamic programming problem given a law of motion for K_t, and (ii) evaluating the likelihood by integrating over the unobserved dependence state. We next show how the Bernstein weights $\{w_{rs}\}$ imply a finite-state Markov transition law for K_t, and then exploit this to write the model as a hidden Markov model.

3.3 Transitions between latent states

The latent indices $\{K_{jt}\}$ provide a discrete state variable that carries all serial dependence. The Bernstein weights have a direct transition interpretation. Since $\mathbb{P}\{K_{j1} = r, K_{j2} = s\} = w_{rs}$ and the Bernstein constraints imply $\mathbb{P}\{K_{j1} = r\} = \sum_{s=1}^m w_{rs} = 1/m$ and $\mathbb{P}\{K_{j2} = s\} = \sum_{r=1}^m w_{rs} = 1/m$, Bayes' rule gives the (time-homogeneous) transition matrix

$$\pi_{rs} \equiv \mathbb{P}\{K_{j2} = s \mid K_{j1} = r\} = \frac{w_{rs}}{1/m} = m\, w_{rs}, \qquad r, s \in \{1, \ldots, m\}. \tag{9}$$

[9] The derivation of the inclusive value follows by defining $Z_{it} = -\log U_{it}$, and rewriting the within-period maximum as

$$\max_{1\leq\ell\leq I} \{\tilde{v}_{\ell t} + \varepsilon_{\ell t}\} = \max_\ell \{-\log(Z_{\ell t} e^{-\tilde{v}_{\ell t}})\} = -\log\left(\min_\ell \{Z_{\ell t} e^{-\tilde{v}_{\ell t}}\}\right).$$

Using the result in Lemma 2 we can then proceed to integrate against the joint density to give the desired result.

By construction, $\sum_{s=1}^{m} \pi_{rs} = \sum_{r=1}^{m} \pi_{rs} = 1$. Intuitively, w_{rs} allocates mass to persistent pairs (r, s): diagonal concentration ($r \approx s$) implies persistence in the latent rank, while diffuse off-diagonal mass implies mobility.

Across alternatives, Assumption SC implies that the latent indices evolve independently. Hence for the latent vector $K_t = (K_{1t}, \dots, K_{It})$ we obtain the product transition

$$\mathbb{P}\{K_2 = (s_1, \dots, s_I) \mid K_1 = (r_1, \dots, r_I)\} = \prod_{i=1}^{I} \pi_{r_i s_i}. \tag{10}$$

These transitions are the only channel through which the copula weights $\{w_{rs}\}$ affect dynamic behaviour: conditional on K_t, period-t CCPs and inclusive values are given by Proposition 1. Assumption CI provides the complementary restriction on the transition of the observed state needed for the dynamic programming problem to condition on (x_t, K_t).

3.4 Hidden Markov model structure and likelihood evaluation

Another practical consequence of the latent-index representation is that it enables efficient likelihood evaluation. Because K_t evolves as a finite-state Markov chain and, conditional on (x_t, K_t), choice probabilities are available in closed form, our serial-dependence copula model can be viewed as a HMM with observed covariates x_t. Conditional on the model parameter vector η, and given a solved policy, define the time-t emission probability

$$e_t^y(k) \equiv \mathbb{P}_\eta\{y_t \mid x_t, K_t = k\}, \qquad k \in \{1, \dots, m\}^I,$$

which is obtained from Proposition 1 (with the value indices $\tilde{v}_t$ evaluated at the solved continuation value). In the general case where the observed state transition depends on the latent dependence state, Assumption CI allows the conditional law of x_{t+1} to depend on contemporaneous shocks only through K_t. For likelihood evaluation it is convenient to treat the time-t observation as the pair (y_t, x_{t+1}) whenever x_{t+1} is observed in the panel, and as y_t in the last observed panel period. For notational simplicity we suppress the individual index and present the recursion for a single history $(x_1, \dots, x_T, y_1, \dots, y_T)$; for an unbalanced panel with individual-specific lengths T_i one applies the same recursion separately for each i up to T_i. Accordingly, define the (time-varying) emission

$$e_t(k) \equiv \begin{cases} \mathbb{P}_\eta\{y_t \mid x_t, K_t = k\}\, p_\eta(x_{t+1} \mid x_t, y_t, K_t = k), & t = 1, \dots, T-1, \\ \mathbb{P}_\eta\{y_T \mid x_T, K_T = k\}, & t = T, \end{cases}$$

and note that if x_{T+1} is observed in a given application, one simply uses the first line also for $t = T$. Operationally, K_t follows a Markov chain with transition law $\mathbb{P}\{K_t \mid K_{t-1}\}$ that does not depend on (x_t, y_t).[10] Let $\alpha_t(k)$ denote the filtered distribution after processing the time-t observation (as defined above), and initialise the prior with $\alpha_{1|0}(k) = m^{-I}$. At $t = 1$ we set $\alpha_{1|0}$ as the predicted distribution and update it using $e_1(\cdot)$. For $t \geq 2$ we apply the standard

[10] Connault (2016) also uses forward filtering for likelihood evaluation. In the Hidden Rust model, the emission probabilities are standard logit, while here they are given by the Bernstein closed forms.

predict/update recursion (Hamilton, 1989):

$$\text{predict:} \quad \alpha_{t|t-1}(k) = \sum_{k'} \mathbb{P}\{K_t = k \mid K_{t-1} = k'\}\, \alpha_{t-1}(k'), \tag{11}$$

$$\text{update:} \quad \alpha_t(k) = \frac{\alpha_{t|t-1}(k)\, e_t(k)}{\sum_{k'} \alpha_{t|t-1}(k')\, e_t(k')}. \tag{12}$$

The period-t contribution to the log-likelihood is

$$\log\Big(\sum_k \alpha_{t|t-1}(k)\, e_t(k)\Big), \tag{13}$$

and the full sample log-likelihood contribution is obtained by summing (13) over $t = 1, \dots, T$. When $p_\eta(x_{t+1} \mid x_t, y_t, K_t)$ is independent of K_t, the x-transition factor in $e_t(k)$ does not affect filtering over K_t and simply contributes the standard observed-state term $\sum_{t=1}^{T-1} \log p_\eta(x_{t+1} \mid x_t, y_t)$. If one wishes to work with the conditional likelihood of choices given the realised state history, then one drops the x-transition factor and uses emissions $e_t^y(k) = \mathbb{P}_\eta\{y_t \mid x_t, K_t = k\}$ for all t.

3.5 Dependence properties of the Bernstein copula

This section records convenient formulas linking the Bernstein weights $\{w_{rs}\}$ to standard dependence measures. These results are useful for interpreting and designing low-dimensional weight parametrisations. Derivations are presented in Appendix A.3.

Proposition 2 (Rank and linear correlations)**.** *Fix an alternative j and let $(U_{j,t-1}, U_{jt})$ follow the Bernstein copula (2) with latent indices $(K_{j,t-1}, K_{jt})$ and weights $\{w_{rs}\}$. Then:*

1. (Spearman rank correlation of $(U_{j,t-1}, U_{jt})$.) *Spearman's rho equals*

$$\rho_S(U_{j,t-1}, U_{jt}) = 12\,\mathbb{E}[U_{j,t-1}U_{jt}] - 3 = \frac{12}{(m+1)^2} \sum_{r=1}^{m} \sum_{s=1}^{m} w_{rs}\, rs \;-\; 3. \tag{14}$$

2. (Pearson correlation of $(\varepsilon_{j,t-1}, \varepsilon_{jt})$.) *Let γ denote the Euler–Mascheroni constant and define*

$$\mu_s \equiv \mathbb{E}[\varepsilon_{jt} \mid K_{jt} = s] = \gamma + \sum_{q=0}^{m-s} \kappa_{sq} \log \lambda_{sq}, \tag{15}$$

where $\{\lambda_{sq}\}$ and $\{\kappa_{sq}\}$ are defined in (5). Since ε_{jt} is marginally T1EV, $\mathbb{E}[\varepsilon_{jt}] = \gamma$ and $\mathrm{Var}(\varepsilon_{jt}) = \pi^2/6$, and therefore

$$\mathrm{Corr}(\varepsilon_{j,t-1}, \varepsilon_{jt}) = \frac{\sum_{r=1}^{m} \sum_{s=1}^{m} w_{rs}\, \mu_r \mu_s - \gamma^2}{\pi^2/6}. \tag{16}$$

Corollary 1 (Maximal and minimal Spearman dependence)**.** *Over all nonnegative weights $\{w_{rs}\}$ satisfying the marginal constraints in (3), Spearman's rho in (14) satisfies*

$$-\frac{m-1}{m+1} \;\le\; \rho_S(U_{j,t-1}, U_{jt}) \;\le\; \frac{m-1}{m+1}.$$

The upper bound is attained by the comonotone (diagonal) weights $w_{rs} = \mathbb{1}\{s = r\}/m$, while the lower bound is attained by the countermonotone (anti-diagonal) weights $w_{rs} = \mathbb{1}\{s = m+1-r\}/m$.

Corollary 2 (Maximal and minimal Pearson dependence)**.** *Over all nonnegative weights $\{w_{rs}\}$ satisfying the marginal constraints in (3), the Pearson correlation in (16) satisfies*

$$\frac{\frac{1}{m}\sum_{s=1}^{m}\mu_s\mu_{m+1-s}-\gamma^2}{\pi^2/6} \le \operatorname{Corr}(\varepsilon_{j,t-1},\varepsilon_{jt}) \le \frac{\frac{1}{m}\sum_{s=1}^{m}\mu_s^2-\gamma^2}{\pi^2/6},$$

where μ_s is defined in (15). The upper and lower bounds are attained by the same comonotone and countermonotone weights as in Corollary 1.

3.6 Low-dimensional weight parametrisations

The Bernstein weights $W = (w_{rs})$ form an $m \times m$ nonnegative array with uniform row and column sums (each equal to $1/m$). This is a flexible dependence specification but it is high-dimensional: after imposing the marginal constraints there are $(m-1)^2$ degrees of freedom.

For interpretation and estimation it is useful to restrict attention to a low-dimensional subfamily that (i) automatically satisfies the Bernstein feasibility constraints and (ii) admits an interpretable dependence index (e.g., persistence in latent ranks). We give two examples. The first is a one-parameter family spanning natural benchmark patterns. The second discretises a target copula family, with the parameter dimension matching that of the target copula.

Example 1: One-parameter family spanning benchmark patterns. Let $w_{rs}^{\text{ind}} \equiv m^{-2}$ denote the independence weights, $w_{rs}^{+} = \mathbb{1}\{s = r\}/m$ the comonotone weights, and $w_{rs}^{-} = \mathbb{1}\{s = m+1-r\}/m$ the countermonotone weights. Each of these weight matrices is feasible: $w_{rs} \ge 0$ and $\sum_s w_{rs} = \sum_r w_{rs} = 1/m$. For $\theta \in [-1,1]$, define

$$w_{rs}(\theta) \equiv (1-|\theta|)\, w_{rs}^{\text{ind}} + \theta_+\, w_{rs}^{+} + (-\theta)_+\, w_{rs}^{-}, \qquad \theta_+ \equiv \max\{\theta, 0\}. \tag{17}$$

Feasibility is immediate because $w_{rs}(\theta)$ is a convex combination of feasible weight matrices.[11] The endpoints $\theta = 0$, $\theta = 1$, and $\theta = -1$ correspond to independence, comonotone matching, and countermonotone matching, respectively. Finally, using (14), the parameter θ maps linearly into Spearman's rho according to

$$\rho_S(U_{j,t-1}, U_{jt}; \theta) = \frac{m-1}{m+1}\,\theta. \tag{18}$$

Thus θ is a convenient dependence index with the same sign and ordering as Spearman's rho, but it is not equal to Spearman's rho at finite m. The induced latent-rank transition matrix is $\pi(\theta) = m\, w(\theta)$: for $\theta > 0$ the mass of π concentrates near the diagonal (persistent ranks), while for $\theta < 0$ it concentrates near the anti-diagonal. Figure 1 in the Appendix illustrates the dependence structure for alternative values of θ at fixed degree.

[11] Note however, that (17) is non-differentiable at $\theta = 0$. There are many ways to smooth this; a natural choice is to use a softmax function.

Example 2: Approximating a target copula. Suppose we have a target copula C^* that we wish to approximate. A canonical way to construct Bernstein weights from C^* is to take the $m \times m$ grid rectangle probabilities

$$w_{rs}(C^*) = C^*\Big(\frac{r}{m}, \frac{s}{m}\Big) - C^*\Big(\frac{r-1}{m}, \frac{s}{m}\Big) - C^*\Big(\frac{r}{m}, \frac{s-1}{m}\Big) + C^*\Big(\frac{r-1}{m}, \frac{s-1}{m}\Big). \tag{19}$$

Because a copula is 2-increasing, the rectangle probabilities in (19) are nonnegative. Moreover, the copula marginal constraints imply uniform row and column sums, so the weights automatically satisfy the Bernstein feasibility constraints. As $m \to \infty$ the resulting Bernstein copula converges uniformly to C^* (Sancetta and Satchell, 2004). If $C^*(\cdot, \cdot; \boldsymbol{\theta})$ is a parametric copula family, applying (19) yields a low-dimensional weight model indexed by $\boldsymbol{\theta}$ (with scalar θ as a special case). Figure 2 in the Appendix illustrates the approximation of a Gaussian copula calculated using (19) for alternative values of m.[12]

Although the full Bernstein class is exactly nested across degrees via Bernstein degree elevation, a *restricted* low-dimensional subclass defined by a fixed functional form for $W = (w_{rs})$ at each degree is generally not closed under degree elevation. In particular, if one changes the degree from m to $m+1$, the degree-elevated weights need not lie in the same restricted subclass. In applications one therefore typically either fixes m when estimating a low-dimensional subclass, or constructs a projective low-dimensional family explicitly (e.g., by degree-elevating a parametric weight table defined at a base degree).

3.7 Extensions and implementation remarks

The baseline structure is convenient for exposition: for each alternative, a bivariate Bernstein copula links $(U_{j,t-1}, U_{jt})$ over time while preserving Type-I extreme value marginals, so that conditional on the latent indices the shocks are independent Beta draws and the period CCPs and inclusive values admit finite-sum closed forms. The same conditioning logic extends to a range of useful generalisations. We briefly highlight three: alternative-specific copulas (useful for parsimony), across-alternative dependence (to relax within-period independence and IIA), and higher-order serial dependence (to allow longer memory).

Alternative-specific copulas. A specific class of structured dependence arises when each alternative has its own Bernstein copula. This is useful when the researcher wants to focus on serial dependence in some alternatives but not others, as it reduces the size of the state space. Suppose that each alternative j has its own degree m_j copula C_j with weights $\{w_{rs}^{(j)}\}$ and latent indices $\{K_{jt}\}$. Then, conditional on the full latent state $K_t = (K_{1t}, \ldots, K_{Jt})$, the period-$t$ CCPs and inclusive values continue to be given by Proposition 1, but with alternative-specific

[12] The degree m controls the resolution of the latent rank state $K_t \in \{1, \ldots, m\}$ and therefore the accuracy with which the copula-induced conditional law of shocks is represented. When one estimates an unrestricted weight table W, increasing m enlarges the parameter space. Here, increasing m does not add free parameters, but it still changes the dependence structure that can be represented and increases computational burden. The choice of m is therefore analogous to selecting the number of grid points in a discretised dynamic program, the number of terms in a series approximation, or the number of quadrature nodes. One chooses m large enough that discretisation error is negligible for the objects of interest, but not so large that computation and numerical instability dominate.

weights, coefficients, and summation limits. The latent-index transitions are now governed by the alternative-specific transition matrices $\{\pi_{rs}^{(j)} = m_j w_{rs}^{(j)}\}$, and the full latent vector K_t evolves according to the product transition.

Across-alternative (within-period) dependence. Across-alternative (within-period) dependence matters because it directly shapes substitution patterns: it relaxes the independence structure that underlies the i.i.d.-logit IIA property.[13] Such dependence can be introduced by considering multivariate Bernstein copulas over $(U_{1t}, \ldots, U_{Jt})$ at each time t. The key is that the Bernstein copula structure continues to yield Beta marginals for each U_{jt} conditional on the latent indices $\{K_{jt}\}$, so the selection conditioning arguments continue to apply. The main cost is computational: the multivariate Bernstein weights form an m^J-dimensional tensor, so naive evaluation scales exponentially in the number of alternatives. Appendix C states the static formulas and discusses computational considerations and alternative copula classes.

Higher-order serial dependence. First-order serial dependence captures persistence in unobserved tastes or costs with minimal state augmentation. However, if the unobserved utility component reflects slow-moving latent tastes, switching costs, or habit/loyalty, the implied dependence may have longer memory than a one-lag structure. The same copula approach extends to dependence of order $L > 1$, where the conditional law of $\varepsilon_{j,t+1}$ depends on the last L shocks $(\varepsilon_{jt}, \varepsilon_{j,t-1}, \ldots, \varepsilon_{j,t-L+1})$.[14]

Operationally, we replace the bivariate Bernstein copula for $(U_{jt}, U_{j,t+1})$ with an $(L+1)$-variate Bernstein copula for the block $(U_{j,t-L+1}, \ldots, U_{j,t+1})$:

$$C(u_1, \ldots, u_{L+1}) = \sum_{k_1=1}^{m} \cdots \sum_{k_{L+1}=1}^{m} w_{k_1 \cdots k_{L+1}} \prod_{\ell=1}^{L+1} B_{k_\ell, m}(u_\ell),$$

with the usual uniform marginal constraints for each coordinate. The associated latent indices are $(K_{j,t-L+1}, \ldots, K_{j,t+1}) \in \{1, \ldots, m\}^{L+1}$. A convenient Markov state at time t is the length-L vector of the most recent indices, $S_{jt} \equiv (K_{j,t-L+1}, \ldots, K_{jt}) \in \{1, \ldots, m\}^L$, which evolves by a shift-and-append update: draw $K_{j,t+1}$ from the conditional distribution implied by the $(L+1)$-variate weights given S_{jt}, and set $S_{j,t+1} = (K_{j,t-L+2}, \ldots, K_{j,t+1})$.

Crucially, conditional on the within-period latent vector $K_t = (K_{1t}, \ldots, K_{Jt})$, the period-$t$ CCPs and inclusive values retain the same finite-sum closed forms as in the first-order case. Higher-order dependence enters only through the law of motion for the latent state in the dynamic recursion. The tradeoff is computational: the latent state space for each alternative grows from size m to m^L, so L must typically be kept small or accompanied by further structure/restrictions.

[13]The well-known "red bus/blue bus" paradox (McFadden, 1974) arises precisely because the i.i.d.-logit model assumes within-period independence across alternatives (equivalently, the errors are independent and hence satisfy IIA). Allowing within-period dependence across alternatives is one way to break this restriction and obtain more reasonable substitution patterns. This example is developed further in Appendix C.

[14]Multi-period coherence is obtained by viewing an Lth-order process as a first-order Markov process on the augmented state (the most recent L latent indices).

4 Illustrative example: occupational choice

This section illustrates the implications of serial dependence in a simple dynamic occupational choice model with occupation specific human capital as in, e.g., Keane and Wolpin (1997) and Kambourov and Manovskii (2009). For expositional simplicity, we present the model for $T = 2$ periods, but in the subsequent Monte Carlo experiment we extend this to panels of arbitrary length.

The agent chooses an occupation $y_t \in \{1,2\}$ in each period $t \in \{1,2\}$ to maximise expected discounted utility. Let the deterministic baseline log-wage index in occupation j be ω_j. There are two sources of persistence in this model. Firstly, there is state dependence: choosing occupation j in period 1 builds occupation-specific human capital that raises period-2 wages in that occupation by $h > 0$.[15] Second, there is serial dependence in the unobservables, with the stochastic component satisfying Assumptions CL, SC, and BC. Period utility is additively separable in deterministic (log-wage) and stochastic components such that

$$v_{j1} = \omega_j + \varepsilon_{j1}, \tag{20a}$$
$$v_{j2} = \omega_j + h\,\mathbb{1}\{y_1 = j\} + \varepsilon_{j2}. \tag{20b}$$

The model is solved by backward induction, exploiting the implied latent-state structure. Let $K_t \equiv (K_{1t}, K_{2t}) \in \{1,\dots,m\}^2$ denote the pair of latent indices across the two occupations in period t. At $t = 2$ the agent is myopic and maximises (20b). The expected maximum utility for period 2, conditional on the latent state K_2 and period-1 choice y_1, can be obtained in closed form using (8). Denote this $\widetilde{G}(K_2, y_1)$.

At $t = 1$ the agent chooses y_1 to maximise current utility plus the discounted expected period-2 utility. Conditional on the observed latent state K_1, this problem involves a continuation value that integrates over K_2. Let π denote the $m \times m$ transition for each occupation's latent index with $\pi_{rs} = \mathbb{P}\{K_{j2} = s \mid K_{j1} = r\} = m\,w_{rs}$, and let $\Pi \equiv \pi \otimes \pi$ denote the induced transition matrix for K_t. The agent solves

$$y_1 \in \arg\max_{j\in\{1,2\}} \Big\{ \omega_j + \varepsilon_{j1} + \beta \sum_{K_2} \Pi(K_2 \mid K_1)\, \widetilde{G}\big(K_2, j\big) \Big\}.$$

Again, the period-1 choice probabilities and inclusive values conditional on K_1 are available in closed form from Proposition 1.

In this illustration we set $\omega_1 = 0$, $\omega_2 = 0.2$, $h = 0.4$, and $\beta = 0.95$. Table 1 reports the impact of serial dependence on occupational choices and switching behaviour over time. We use the one-parameter weight family in (17) to summarise serial dependence with a single parameter θ that maps linearly into Spearman's rank correlation. The table presents results for alternative values of θ at fixed degree $m = 5$. It shows that as θ increases (more persistence in the latent ranks), current "comparative advantage" tends to carry over to the next period, so agents are more inclined to stick with the same occupation and switching probabilities fall. Similarly, when future shocks are more aligned with current shocks, the dominance from the initial wage gap ($\omega_2 > \omega_1$) is eroded, which pushes the share $\mathbb{P}\{y_1 = 1\}$ toward 1/2.

[15] In the $T > 2$ generalisation, the wages in occupation j at time $t > 1$ are given by $\omega_j + h \sum_{s=1}^{t-1} \mathbb{1}\{y_s = j\}$.

Dependence θ	$\mathbb{P}\{y_1 = 1\}$	$\mathbb{P}\{y_2 = 1\}$	$\mathbb{P}\{y_2 \neq y_1\}$
-0.80	0.431	0.430	0.553
0.00	0.441	0.441	0.397
0.80	0.446	0.446	0.209

Table 1: Occupational choices and switching under serial dependence. Moments calculated at degree $m = 5$ for alternative values of the one-parameter weight θ in (17). Entries report unconditional choice probabilities $\mathbb{P}\{y_1 = 1\}$, $\mathbb{P}\{y_2 = 1\}$, and the switching rate $\mathbb{P}\{y_2 \neq y_1\}$.

Identification (sketch). We discuss identification of our serial-copula model more generally in Appendix D. Here we explain how identification in this occupational-choice example follows from the arguments in Hu and Shum (2012).

Step 1: reduced-form HMM identification from short panels. Conditional on observed states (experience/human-capital and potentially other observables), the choice process admits a finite-state HMM representation: the latent state is the rank index $K_t \in \{1, \dots, m\}^2$, its transition is $\Pi(\theta)$, and the emission probabilities are the choice probabilities $\mathbb{P}\{y_t = 1 \mid x_t, K_t\}$ implied by the T1EV/ Bernstein-mixture structure. Identification of the reduced-form HMM objects from short panels is the first step in Hu and Shum (2012). Their sufficient conditions are stated for a broad class of models and, under stationarity, typically require at least four periods to construct the matrix factorizations used in their argument. By contrast, in the finite-state HMM literature, three consecutive observations can suffice under additional full-rank conditions (Hsu et al., 2012). In either case, the reduced-form HMM parameters are identified up to a permutation of latent-state labels.

Step 2: fixing labels using the rank interpretation. In a generic HMM, relabelling latent states is observationally irrelevant. In our model the latent states correspond to ordered rank regions, so there is an economically meaningful labelling: higher K_{jt} corresponds to higher quantile regions of ε_{jt}. This ordering pins down the permutation and aligns the reduced-form HMM states with the structural rank indices.

Step 3: mapping HMM objects to structural parameters. Given the structural labelling, the emission probabilities equal the closed-form logit/Bernstein-mixture CCPs from (7) with deterministic indices determined by (ω_2, h) and observed experience. Variation in observed states (accumulated experience) shifts these deterministic indices, which allows (ω_2, h) to be recovered from the labelled emission probabilities. The transition matrix is restricted to the low-dimensional copula family $\Pi(\theta)$, so θ is recovered by inverting this map.

4.1 Simulation experiment

To illustrate finite-sample performance, we conduct a small Monte Carlo experiment under short panels. As above, we fix $\beta = 0.95$, $\omega_1 = 0$, and degree $m = 5$. The parameters that we estimate have the data-generating values $\omega_2 = 0.2$ and $h = 0.4$, with three values of the persistence parameter considered, $\theta \in \{-0.8, 0, 0.8\}$, corresponding to negative, zero, and positive serial dependence. For each parameter configuration, we simulate 1000 datasets of size $n \in \{100, 1000\}$, and panel dimension $T \in \{3, 5, 10\}$. Estimation is performed by maximum likelihood using the hidden Markov recursion to integrate out the latent dependence states, as described in Section 3.4.

Likelihood. Let the observed data for individual $i \in \{1, \dots, n\}$ be $(y_{i1}, \dots, y_{iT}) \in \{1, 2\}^T$. Assume that individuals are observed from period 1, such that experience x_t is a deterministic function of observed past choices $(y_{i1}, \dots, y_{i,t-1})$. The latent dependence state is the pair of latent ranks across occupations, $K_{it} = (K_{1it}, K_{2it}) \in \{1, \dots, m\}^2$. Under the Bernstein copula, the induced Markov transition is given by $\Pi(\theta) \equiv \pi(\theta) \otimes \pi(\theta)$ with $\pi_{rs}(\theta) = m\, w_{rs}(\theta)$. We initialise with $\alpha_{i1|0}(k) = m^{-2}$.

Given parameters $\eta = (\theta, \omega_2, h)$, the model solution delivers closed-form CCPs. In this experiment, the observed state x_{it} (experience) is a deterministic function of past choices $(y_{i1}, \dots, y_{i,t-1})$, and its evolution does not enter the likelihood. The individual likelihood contribution can then be computed efficiently using the standard HMM predict/update recursion with emission probabilities $e_{it}(k) \equiv p_\eta(y_{it} \mid x_{it}, k)$.

Results. Table 2 reports results for the correctly specified Bernstein-copula model. As expected, increasing either the cross-sectional sample size n or the panel length T sharpens identification and reduces sampling error across all designs. A notable pattern is that estimates are typically most precise under negative dependence ($\theta < 0$). Intuitively, negative serial dependence makes the latent shocks less persistent, generating more switching and richer within-agent variation, which helps disentangle true state dependence (h) from persistence in unobserved heterogeneity over time.[16] Overall, the results illustrate that the Bernstein-copula framework can be used to estimate dynamic discrete choice models with serial dependence in unobserved shocks.[17]

To quantify the consequences of ignoring serial dependence, we also estimate each simulated dataset under the misspecified restriction $\theta = 0$ (equivalently, $m = 1$ so that shocks are i.i.d. and CCPs reduce to the standard dynamic logit form). Results are presented in Table E.1 in the Appendix. A clear pattern emerges. When the true dependence is positive ($\theta > 0$), the i.i.d.-logit fit attributes excess persistence in choices to stronger state dependence, leading to upward bias in $\hat{h}$ and an offsetting downward bias in $\hat{\omega}_2$. When the true dependence is negative ($\theta < 0$), the i.i.d.-logit fit instead lowers $\hat{h}$ to facilitate switching and compensates by raising $\hat{\omega}_2$. As the panel length T increases, the experience parameter gets pinned down more tightly by experience variation (i.e. the bias in $\hat{h}$ falls), with the remaining misspecification manifesting in the estimated level parameter $\hat{\omega}_2$.

[16] In some of our specifications, we estimate the model with $T = 3$. This does not contradict Hu and Shum (2012). Their panel-length conditions are sufficient for identification in a much more general setting (and under stationarity are typically stated with $T \geq 4$), whereas here the combination of discrete latent ranks, T1EV/mixture emissions, and the low-dimensional copula parameterization strengthen identification in short panels.

[17] The serial copula model is a restricted member of the Arcidiacono and Miller (2011) class of models with discrete Markov latent heterogeneity. Both frameworks induce a hidden Markov model on observed choices, with the copula's latent rank indices $K_t \in \{1, \dots, m\}^J$ playing the role of Arcidiacono and Miller's unobserved state s_t. The structural restriction is that the transitions are governed by $\Pi(\theta)$ and the emissions are given by the Bernstein-mixture CCPs conditional on (x_t, K_t). A generic EM/CCP approach identifies the reduced-form HMM objects only up to a permutation of latent labels. However, in the present model the latent states have a rank interpretation that induces a natural ordering, so (in principle) one can fix the permutation by imposing an ordering restriction or by matching the reduced-form HMM states to the structural rank bins. Implementing such a label-fixing step in a numerically robust way is nontrivial, so we focus on maximum likelihood estimators.

n	T	Positive dep. ($\theta = 0.8$)			Independence ($\theta = 0.0$)			Negative dep. ($\theta = -0.8$)		
		$\hat{\theta}$	$\hat{\omega}_2$	$\hat{h}$	$\hat{\theta}$	$\hat{\omega}_2$	$\hat{h}$	$\hat{\theta}$	$\hat{\omega}_2$	$\hat{h}$
100	3	0.73	0.18	0.48	0.06	0.21	0.34	−0.79	0.20	0.39
		(0.28)	(0.15)	(0.31)	(0.30)	(0.12)	(0.20)	(0.12)	(0.08)	(0.08)
100	5	0.78	0.19	0.42	0.01	0.21	0.39	−0.79	0.21	0.40
		(0.13)	(0.13)	(0.13)	(0.16)	(0.07)	(0.05)	(0.07)	(0.05)	(0.04)
100	10	0.79	0.20	0.39	−0.02	0.21	0.40	−0.79	0.20	0.40
		(0.08)	(0.11)	(0.06)	(0.18)	(0.04)	(0.02)	(0.07)	(0.03)	(0.02)
1000	3	0.77	0.19	0.44	0.01	0.20	0.39	−0.80	0.20	0.40
		(0.12)	(0.05)	(0.13)	(0.09)	(0.03)	(0.06)	(0.04)	(0.02)	(0.02)
1000	5	0.78	0.19	0.41	0.00	0.20	0.40	−0.80	0.20	0.40
		(0.06)	(0.04)	(0.06)	(0.05)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
1000	10	0.80	0.20	0.40	−0.01	0.20	0.40	−0.80	0.20	0.40
		(0.03)	(0.04)	(0.02)	(0.06)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)

Table 2: Monte Carlo results for occupational choice model with serial dependence (Bernstein copula, $m = 5$). Data generated with parameters $(\omega_2, h) = (0.2, 0.4)$ and varying θ. Each cell reports the mean estimate and average standard error over 1000 simulations. Here n is the cross-sectional sample size and T is the panel length.

5 Empirical exercise: Rust (1987)

This section applies the proposed serial-copula framework to the canonical bus engine replacement model of Rust (1987), which remains a standard benchmark for dynamic discrete-choice methods and estimators. While the model and its solution are well-documented in the literature, for completeness we provide a brief description of the framework and our specific adaptation of it below.

5.1 Baseline model

In the bus engine replacement model of Rust (1987), a fleet manager decides each period whether to replace the engine of a bus based on its observed mileage (odometer reading) x_t since last replacement and unobserved preference shocks. Per-period utility is additively separable,

$$u(x_t, j) + \varepsilon_{jt} = \begin{cases} -\eta_{11}\, x_t/1000 + \varepsilon_{1t} & \text{if } j = 1, \\ -RC + \varepsilon_{2t} & \text{if } j = 2, \end{cases} \tag{21}$$

where $j \in \{1, 2\}$ is the decision to either keep ($j = 1$) or replace ($j = 2$) the engine, and ε_{jt} is a choice-specific shock observed by the agent but not the econometrician. The parameter $RC > 0$ is the fixed replacement cost, with replacement resetting the mileage to zero. If keeping the engine, a linear-in-miles maintenance cost applies, with $\eta_{11} > 0$ governing how maintenance costs rise with mileage.

The odometer reading is discretized into 90 intervals, $X = \{1, \ldots, 90\}$, whose mileage evolution is stochastic as governed by the sparse state transition matrix

$$p(x_{t+1} \mid x_t, j; \eta_3) = \begin{cases} \eta_{30} & x_{t+1} = x_t, \\ \eta_{31} & x_{t+1} = x_t + 1, \\ \eta_{32} & x_{t+1} = x_t + 2, \end{cases}$$

with the non-negative parameters satisfying $\eta_{30} + \eta_{31} + \eta_{32} = 1$. Finally, the environment is stationary, with an infinite horizon, and with the continuation value discounted at rate β. Omitting time subscripts, the agent's value function can be formulated as a dynamic programming problem

$$V(x, \varepsilon) = \max_{j \in \{1,2\}} \Big\{ u(x, j) + \varepsilon_j + \beta\, \mathbb{E}[V(x') \mid x, j] \Big\},$$

where the expectation is over the next period's state x' and shocks ε'. Rust (1987) assumed that the shocks are i.i.d. over time and follow a Type-I extreme value distribution marginally.

5.2 Data and baseline model

We use the original bus-engine data from Rust (1987), which tracks many buses at a monthly frequency and records both the odometer reading and whether the engine was replaced in each period. The sample is partitioned into four bus groups according to observable characteristics. See Rust (1987) for full details on the data construction and grouping.

We focus on the samples corresponding to Rust's Table IX: groups (1–3), group (4), and groups (1–4). The model is estimated separately on each of these three samples to facilitate comparison with Rust's original results, and we set $\beta \in \{0.9999, 0\}$, matching Rust's reported discount-factor sensitivity checks. A nested fixed-point approach is used to estimate the model parameters. We solve the fixed point problem by first iterating on the Bellman equation, and then switching to Newton-Kantorovich iterations to achieve rapid convergence (Rust, 1987; Iskhakov et al., 2016).[18] Estimation proceeds by maximum likelihood, with the transition parameters (η_{30}, η_{31}) estimated in a first step.

Results are reported in the first three columns of Table 3 for the baseline i.i.d.-shock model, with the estimates very closely matching Rust's reported results.

5.3 Serial-dependence model

We retain the additive structure presented in (21) and T1EV marginals, but as in Norets (2009) and Reich (2018), we allow the unobserved engine-specific maintenance costs to be serially dependent over time.[19] We implement this by applying the serial Bernstein copula

[18] Su and Judd (2012) emphasise the importance of exploiting sparsity in the transition structure of Rust's model in the context of a mathematical programming with equilibrium constraints (MPEC) estimation approach. We also exploit this structure in our implementation. The Newton-Kantorovich requires the solution of large linear systems at each iteration, which we solve using a sparse linear-algebra solver. This allows us to solve and estimate the model efficiently even with a large number of latent states, as in the following.

[19] We maintain that the replacement-option shock is i.i.d. This is a special case of the alternative-specific copula extension in Section 3.7, with a nontrivial degree m copula for $j = 1$ and $m = 1$ (independence) for $j = 2$.

(Assumptions SC and BC), which introduces a single discrete latent index $K_{1t} \in \{1, \ldots, m\}$, that evolves as a Markov chain with a transition matrix that is derived from the Bernstein weights via (9).

To maintain parsimony, and also to achieve closer comparability with Norets (2009) and Reich (2018), we construct the Bernstein weights to approximate a target Gaussian copula with correlation coefficient θ. Thus, the second step of estimation in our serial-dependence model involves the augmented parameter vector $\eta = (RC, \eta_{11}, \theta)$. Conditional on the observed state x_t and latent index K_{1t}, both the choice probability $p_\eta(j_t \mid x_t, K_{1t})$ and the corresponding inclusive value are available in closed form from Proposition 1. In the infinite-horizon problem these enter the Bellman fixed point. A convenient feature of the serial-copula specification is that the infinite-horizon Bellman operator remains a contraction on the augmented state space (x_t, K_{1t}), ensuring existence and uniqueness of the fixed point and convergence of value function iteration (see Appendix A.4). Given a solved policy, the likelihood integrates out the latent Markov state using the forward recursion described in Section 3.3.

5.4 Serial-dependence results

In Table 3 we additionally report results obtained for both discount factors $\beta \in \{0.9999, 0\}$ and for Bernstein copula degrees $m \in \{4, 8\}$, alongside the baseline i.i.d.-shock model ($m = 1$). Across all samples and for both discount factors, we obtain an estimated persistence parameter θ that is both positive and sizable. The estimated magnitude of the parameter, and the improvement in the log-likelihood, is larger for the lower discount factor. This is consistent with the idea that when continuation values are shut down, serial correlation in shocks plays a relatively larger role in explaining persistence in replacement behaviour.

In the $m = 8$ case when $\beta = 0.9999$, the estimated value of θ implies a Spearman rank correlation of between 0.67 and 0.70 (depending on the sample) for shocks one period apart. This indicates substantial persistence in unobserved maintenance costs, although the estimates are somewhat imprecise. The dependence is stronger than is estimated when $m = 4$ (where rank correlations average around 0.53), but further changes in the implied dependence when higher copula degrees are considered are more muted (see Appendix E). The estimated persistence is much stronger when $\beta = 0$, where the estimates are on the boundary. They imply a rank correlation of 0.6 when $m = 4$ and 0.78 when $m = 8$. Results stabilise for higher m, with rank correlations exceeding 0.8. Additional results are presented in the Appendix.

Comparison to related findings. A small number of other papers have also extended Rust's model to incorporate serial dependence in unobserved shocks. Norets (2009) introduce an additive AR(1) process in the per-period unobserved utility component, and propose a Bayesian framework with Markov chain Monte Carlo methods to handle the high dimensional integration in the likelihood function. They find limited evidence for serial correlation in the original Rust data, with posterior means near zero and only moderate posterior probability that the persistence parameter is positive. Reich (2018) considers a similar specification, although the implementation is different: a quadrature-based method is used to decompose the high-dimensional integrals and recursively approximate them using numerical quadrature and

	Independent			Copula ($m = 4$)			Copula ($m = 8$)		
	(1-3)	(4)	(1-4)	(1-3)	(4)	(1-4)	(1-3)	(4)	(1-4)
(a) Discount factor $\beta = 0.9999$									
RC	11.727	10.075	9.756	12.012	10.324	10.038	12.109	10.401	10.139
	(1.649)	(1.329)	(0.717)	(2.046)	(1.554)	(0.963)	(2.098)	(1.664)	(0.986)
η_{11}	4.826	2.293	2.628	5.058	2.441	2.827	5.137	2.486	2.900
	(1.190)	(0.518)	(0.384)	(1.496)	(0.732)	(0.524)	(1.550)	(0.820)	(0.545)
η_{30}	0.301	0.392	0.349	0.301	0.392	0.349	0.301	0.392	0.349
	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)
η_{31}	0.688	0.595	0.639	0.688	0.595	0.639	0.688	0.595	0.639
	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)
θ	–	–	–	0.919	0.949	0.956	0.882	0.910	0.925
				(0.380)	(0.406)	(0.126)	(0.365)	(0.444)	(0.127)
$\log \mathcal{L}$	-2713.4	-3304.2	-6059.8	-2713.3	-3304.1	-6059.5	-2713.3	-3304.1	-6059.4
(b) Discount factor $\beta = 0.0000$									
RC	8.299	7.636	7.306	8.715	8.346	7.975	8.853	8.429	8.199
	(0.568)	(0.505)	(0.293)	(0.793)	(0.675)	(0.433)	(0.812)	(0.821)	(0.463)
η_{11}	109.905	71.513	70.277	125.869	93.055	93.196	131.127	95.817	100.930
	(13.628)	(9.088)	(5.684)	(20.018)	(13.344)	(9.386)	(20.651)	(18.807)	(10.333)
η_{30}	0.301	0.392	0.349	0.301	0.392	0.349	0.301	0.392	0.349
	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)
η_{31}	0.688	0.595	0.639	0.688	0.595	0.639	0.688	0.595	0.639
	(0.008)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)	(0.005)
θ	–	–	–	1.000	1.000	1.000	1.000	0.998	1.000
				(0.000)	(0.000)	(0.000)	(0.000)	(0.010)	(0.000)
$\log \mathcal{L}$	-2715.7	-3306.0	-6066.2	-2715.3	-3305.4	-6063.5	-2715.3	-3305.4	-6063.1

Table 3: Estimation results for the Rust (1987) bus engine replacement model. Panel (a) presents results with the monthly discount factor $\beta = 0.9999$ for different bus group samples, and alternative Bernstein copula degree values: independent ($m = 1$), $m = 4$, and $m = 8$. Panel (b) presents results with the monthly discount factor $\beta = 0$. The parameter θ is the correlation coefficient of the target Gaussian copula. Standard errors are in parentheses.

interpolation. Similar to the findings here, the estimated persistence parameter is positive and sizable, but imprecise. Reich (2018) also finds that the parameter estimates vary substantially compared with the i.i.d.-shock model, which is not the case in our results. Finally, using sequential Monte Carlo methods, Blevins (2016) finds strong evidence of persistence in the Rust data (especially for bus group 4), although the source is different: a latent state directly enters the per-period payoffs as well as the state variable transition function.

6 Conclusion

Copula methods provide a convenient way to model dependence separately from marginal distributions, but they have seen limited use in dynamic discrete choice because serially dependent shocks typically destroy the logit closed forms that make Rust-style models easy to solve and estimate.

This paper develops a novel copula-based approach to serial dependence that preserves the familiar logit building blocks. With Type-I extreme value marginals and a Bernstein-polynomial copula, persistence can be summarized by an endogenous finite-state latent rank process. Conditional on the latent rank state, conditional choice probabilities and inclusive values remain available in closed form as finite mixtures of logit terms, so the dynamic program can be solved on an augmented state space without simulation. Estimation is likewise exact: the likelihood integrates out the latent state by standard hidden Markov model filtering, with the latent transition disciplined by copula validity and by the low-dimensional dependence parameterization.

We illustrate the framework in a simulation study of an occupational choice model and in an empirical application to the canonical bus engine replacement problem of Rust (1987). In both settings, allowing for serial dependence provides a tractable way to distinguish persistence due to observed state dynamics from persistence due to unobserved shocks.

An important direction for future work is to connect this approach more directly to the Arcidiacono and Miller (2011) framework for dynamic discrete choice with a discrete latent Markov state estimated by EM/CCP methods. That framework offers a flexible way to accommodate serial dependence (or other omitted persistent factors) by allowing an unobserved regime to shift payoffs and evolve over time. Our copula-based model can be viewed as a structured member of this class: the latent state has a rank interpretation and its transition is restricted to a low-dimensional copula family. This suggests hybrid estimators that use an Arcidiacono–Miller style first stage to recover a flexible latent-state representation, and then impose economically motivated structure to interpret the latent state as persistent shocks rather than arbitrary utility regimes.

References

Aguirregabiria, Victor and Pedro Mira (2002) "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," *Econometrica*, 70 (4), 1519–1543.

——— (2010) "Dynamic discrete choice structural models: A survey," *Journal of Econometrics*, 156 (1), 38–67.

Allman, Elizabeth S., Catherine Matias, and John A. Rhodes (2009) "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 37 (6A), 3099–3132.

Arcidiacono, Peter and Paul B. Ellickson (2011) "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, 3 (Volume 3, 2011), 363–394.

Arcidiacono, Peter and Robert A. Miller (2011) "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity," *Econometrica*, 79 (6), 1823–1867.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75 (5), 1331–1370.

Bierens, Herman J. (2008) "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results," *Econometric Theory*, 24 (3), 749–794.

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

Chiong, Khai Xiang, Alfred Galichon, and Matt Shum (2016) "Duality in Dynamic Discrete-Choice Models," *Quantitative Economics*, 7 (1), 83–115.

Connault, Benjamin (2016) "Hidden Rust Models," technical report, University of Pennsylvania, Department of Economics.

Darsow, William F., Bao Nguyen, and Elwood T. Olsen (1992) "Copulas and Markov Processes," *Illinois Journal of Mathematics*, 36 (4), 600–642.

Eckstein, Zvi and Kenneth I. Wolpin (1989) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey," *The Journal of Human Resources*, 24 (4), 562.

Erdem, Tülin and Michael P. Keane (1996) "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science*, 15 (1), 1–20.

Ericson, Richard and Ariel Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62 (1), 53–82.

Fan, Yanqin and Andrew J Patton (2014) "Copulas in Econometrics," *Annual Review of Economics*, 6 (1), 179–200.

Farlie, D. J. G. (1960) "The performance of some correlation coefficients for a general bivariate distribution," *Biometrika*, 47 (3-4), 307–323.

Galichon, Alfred (2016) *Optimal Transport Methods in Economics*, Princeton, NJ: Princeton University Press.

Gumbel, E. J. (1958) *Statistics of Extremes*, New York: Columbia University.

Hamilton, James D. (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57 (2), 357–384.

Heckman, James J. (1981) "Heterogeneity and State Dependence," in Sherwin Rosen ed. *Studies in Labor Markets*, 91–139, Chicago: University of Chicago Press.

Hotz, V. Joseph and Robert A. Miller (1993) "Conditional Choice Probabilities and the Estimation of Dynamic Models," *The Review of Economic Studies*, 60 (3), 497–529.

Hsu, Daniel, Sham M. Kakade, and Tong Zhang (2012) "A spectral algorithm for learning Hidden Markov Models," *Journal of Computer and System Sciences*, 78 (5), 1460–1480, JCSS Special Issue: Cloud Computing 2011.

Hu, Yingyao and Matthew Shum (2012) "Nonparametric identification of dynamic models with unobserved state variables," *Journal of Econometrics*, 171 (1), 32–44.

Hyslop, Dean R. (1999) "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67 (6), 1255–1294.

Iskhakov, Fedor, Jinhyuk Lee, John Rust, Bertel Schjerning, and Kyoungwon Seo (2016) "Comment On "Constrained Optimization Approaches To Estimation Of Structural Models"," *Econometrica*, 84 (1), 365–370.

Joe, Harry (2014) *Dependence Modeling with Copulas*, Boca Raton, FL: CRC Press.

Johnson, Norman L. and Samuel Kotz (1975) "On some generalized Farlie-Gumbel-Morgenstern distributions," *Communications in Statistics*, 4 (5), 415–427.

Kambourov, Gueorgui and Iourii Manovskii (2009) "Occupational Mobility and Wage Inequality," *The Review of Economic Studies*, 76 (2), 731–759.

Kasahara, Hiroyuki and Katsumi Shimotsu (2009) "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices," *Econometrica*, 77 (1), 135–175.

Keane, Michael P. and Kenneth I. Wolpin (1997) "The Career Decisions of Young Men," *Journal of Political Economy*, 105 (3), 473–522.

Keane, Michael, Petra Todd, and Kenneth I. Wolpin (2011) "The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications," Vol. 4A, 1st edition, Chap. 04, 331–461: Elsevier.

Kristensen, Dennis, Patrick K. Mogensen, Jong Myun Moon, and Bertel Schjerning (2021) "Solving dynamic discrete choice models using smoothing and sieve methods," *Journal of Econometrics*, 223 (2), 328–360, Annals issue: Implementation of Structural Dynamic Models.

Larsen, Bradley J., Florian Oswald, Gregor Reich, and Dan Wunderli (2012) "A test of the extreme value type I assumption in the bus engine replacement model," *Economics Letters*, 116 (2), 213–216.

Lorentz, G.G. (1986) *Bernstein Polynomials*, AMS Chelsea Publishing Series: Chelsea Publishing Company.

Luce, R. Duncan and Patrick Suppes (1965) "Preference, Utility, and Subjective Probability," in R. D. Luce, R. R. Bush, and E. Galanter eds. *Handbook of Mathematical Psychology*, Vol. 3, 249–410, New York: Wiley.

Magnac, Thierry and David Thesmar (2002) "Identifying Dynamic Discrete Decision Processes," *Econometrica*, 70 (2), 801–816.

McFadden, Daniel (1974) "Conditional Logit Analysis of Qualitative Choice Behavior," in Paul Zarembka ed. *Frontiers in Econometrics*, 105–142: Academic Press.

——— (1978) "Modeling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull eds. *Spatial Interaction Theory and Planning Models*, 75–96: North Holland.

McFadden, Daniel L. (1981) "Econometric Models of Probabilistic Choice," in Charles F. Manski and Daniel L. McFadden eds. *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, MA: MIT Press.

Morgenstern, D. (1956) "Einfache Beispiele zweidimensionaler Verteilungen," *Mitteilingsblatt für Mathematische Statistik*, 8, 234–235.

Nelsen, Roger B. (2006) *An Introduction to Copulas*, Springer Series in Statistics, 2nd edition, New York: Springer.

Norets, A. and X. Tang (2013) "Semiparametric Inference in Dynamic Binary Choice Models," *The Review of Economic Studies*, 81 (3), 1229–1262.

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

Norets, Andriy and Kenichi Shimizu (2024) "Semiparametric Bayesian estimation of dynamic discrete choice models," *Journal of Econometrics*, 238 (2), 105642.

Patton, Andrew J. (2012) "A review of copula models for economic time series," *Journal of Multivariate Analysis*, 110, 4–18, Special Issue on Copula Modeling and Dependence.

Reich, Gregor (2018) "Divide and Conquer: Recursive Likelihood Function Integration for Hidden Markov Models with Continuous Latent Variables," *Operations Research*, 66 (6), 1457–1470.

Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, 55 (5), 999–1033.

——— (1994) "Structural estimation of Markov decision processes," in Robert F. Engle and Daniel L. McFadden eds. *Handbook of Econometrics*, Vol. 4, Chap. 51, 3081–3143: Elsevier.

Rust, John and Christopher Phelan (1997) "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, 65 (4), 781–831.

Sancetta, Alessio and Stephen Satchell (2004) "The Bernstein Copula and Its Applications to Modeling and Approximations of Multivariate Distributions," *Econometric Theory*, 20 (3), 535–562.

Schucany, William R., William C. Parr, and John E. Boyer (1978) "Correlation Structure in Farlie-Gumbel-Morgenstern Distributions," *Biometrika*, 65 (3), 650–653.

Segers, Johan, Masaaki Sibuya, and Hideatsu Tsukahara (2017) "The Empirical Beta Copula," *Journal of Multivariate Analysis*, 155, 35–51.

Sklar, Abe (1959) "Fonctions de répartition à n dimensions et leurs marges," *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.

Stinebrickner, Todd R. (2000) "Serially correlated variables in dynamic, discrete choice models," *Journal of Applied Econometrics*, 15 (6), 595–624.

Su, Che-Lin and Kenneth L. Judd (2012) "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 80 (5), 2213–2230.

Appendices

A Additional Proofs

A.1 Proof of Lemma ??

By the generative representation of the Bernstein copula, for any given alternative $i = 1, \ldots, I$, and conditional upon the associated latent state K_{it}, we have $U_{it} \mid (K_{it} = s) \sim \text{Beta}(s, m - s + 1)$. Defining $Z_{it} \equiv -\log U_{it}$, the density of Z_{it} conditional on $K_{it} = s$ is given by

$$f_{Z_{it}|K_{it}=s}(z) = \frac{1}{\mathrm{B}(s, m-s+1)} e^{-sz}(1-e^{-z})^{m-s}, \tag{A.1}$$

where $\mathrm{B}(s, m-s+1)$ is the Beta function. Performing a binomial expansion of the $(1 - e^{-z})^{m-s}$ term, we have

$$(1-e^{-z})^{m-s} = \sum_{q=0}^{m-s} (-1)^q \binom{m-s}{q} e^{-qz},$$

and therefore (A.1) can be rewritten as

$$f_{Z_{it}|K_{it}=s}(z) = \sum_{q=0}^{m-s} (-1)^q \frac{\binom{m-s}{q}(s+q)}{\mathrm{B}(s, m-s+1)(s+q)} e^{-\lambda_{sq} z} = \sum_{q=0}^{m-s} \kappa_{sq}\, \lambda_{sq}\, e^{-\lambda_{sq} z}, \tag{A.2}$$

where the second equality uses the Beta-function identity $\mathrm{B}(s, m-s+1) = (s-1)!(m-s)!/(m!)$ and the definitions of λ_{sq} and κ_{sq} in (5). Thus, conditional on $K_{it} = s$, the marginal density admits a finite expansion as a signed mixture of exponentials. To obtain the joint density of $Z_t \equiv (Z_{1t}, \ldots, Z_{It})$ conditional on K_t, we combine the mixture representation from (A.2) with the independence across alternatives (Assumption SC). The joint density of Z_t factorizes as

$$f_{Z_t|K_t}(z_1, \ldots, z_I) = \prod_{i=1}^{I} f_{Z_{it}|K_{it}}(z_i) = \sum_{q_1=0}^{m-K_{1t}} \cdots \sum_{q_I=0}^{m-K_{It}} \left[\prod_{i=1}^{I} \kappa_{K_{it},q_i}\right] \left(\prod_{i=1}^{I} \lambda_{K_{it},q_i} e^{-\lambda_{K_{it},q_i} z_i}\right),$$

which is the form presented in (4). The product term on the RHS of this equation corresponds to the joint density of independent exponential random variables with rates λ_{K_{it},q_i}. □

A.2 Simplifying the coefficients in the closed-form expressions

Here we describe a recursive scheme to compute the coefficients $\{\kappa_{sq}\}$ defined in (5). For strictly positive integers m and $s \leq m$ we can simplify the coefficients $\{\kappa_{sq}\}$ using the Beta-function identity $\mathrm{B}(s, m-s+1) = (s-1)!(m-s)!/m!$. Therefore,

$$\kappa_{sq} = (-1)^q \frac{m!}{(s-1)!(m-s)!} \frac{1}{s+q} \frac{(m-s)!}{q!(m-s-q)!} = (-1)^q \binom{m}{s+q} \binom{s+q-1}{q}. \tag{A.3}$$

As the coefficients can be written as a product of integers, they are necessarily integers themselves. In practice, we can apply a simple recursive scheme to compute $\kappa_{s,q+1}$ from κ_{sq}

$$\kappa_{s,0} = \binom{m}{s}, \qquad \kappa_{s,q+1} = -\kappa_{s,q} \frac{m-s-q}{s+q+1} \frac{s+q}{q+1}, \qquad q = 0, 1, \ldots, m-s-1.$$

This recursion is numerically stable for $q = 0, 1, \ldots, m-s$ and avoids computing large factorials (which can overflow even for moderate m).

A.3 Proofs for dependence properties

Proof of Proposition 2. For Spearman's rho we have $\rho_S(U_{j1}, U_{j2}) = \mathrm{Corr}(U_{j1}, U_{j2})$. Since U_{j1} and U_{j2} are uniform, $\mathbb{E}[U_{j1}] = \mathbb{E}[U_{j2}] = 1/2$ and hence $\rho_S = 12\,\mathrm{Cov}(U_{j1}, U_{j2}) = 12\,\mathbb{E}[U_{j1}U_{j2}] - 3$. Under the latent-index representation, conditional on $(K_{j1}, K_{j2}) = (r, s)$ we have independence with $\mathbb{E}[U_{j1} \mid K_{j1} = r] = r/(m+1)$, $\mathbb{E}[U_{j2} \mid K_{j2} = s] = s/(m+1)$, so $\mathbb{E}[U_{j1}U_{j2}] = \sum_{r,s} w_{rs}\,(r/(m+1))(s/(m+1))$, which gives (14).

For the Pearson correlation, conditional on $(K_{j1}, K_{j2}) = (r, s)$ the shocks $(\varepsilon_{j1}, \varepsilon_{j2})$ are independent, so $\mathbb{E}[\varepsilon_{j1}\varepsilon_{j2} \mid r, s] = \mathbb{E}[\varepsilon_{j1} \mid r]\,\mathbb{E}[\varepsilon_{j2} \mid s] = \mu_r\mu_s$. Therefore $\mathbb{E}[\varepsilon_{j1}\varepsilon_{j2}] = \sum_{r,s} w_{rs}\mu_r\mu_s$ and $\mathrm{Cov}(\varepsilon_{j1}, \varepsilon_{j2}) = \sum_{r,s} w_{rs}\mu_r\mu_s - \gamma^2$, while the marginal variance is $\pi^2/6$ under T1EV. This yields (16). The identity (15) follows from the exponential-mixture expansion: for $W \sim \mathrm{Exp}(\lambda)$, $-\log W$ has mean $\gamma + \log\lambda$, and linearity applies term-by-term in (5). □

Proof of Corollary 1. By (14), maximizing (or minimizing) $\rho_S(U_{j1}, U_{j2})$ is equivalent to maximizing (or minimizing) the bilinear surplus $\sum_{r,s} w_{rs}\,rs$ over the set of couplings $\{w_{rs}\}$ with uniform marginals. This is a discrete optimal transport problem between two identical uniform measures on $\{1, \ldots, m\}$. The surplus function is rs which is strictly supermodular. Hence the surplus-maximizing coupling is the comonotone (identity) matching, while the surplus-minimizing coupling is the countermonotone matching (Galichon, 2016, Proposition 2.3). Substituting these weights into (14) gives the stated bounds. □

Proof of Corollary 2. By the same argument as in Corollary 1, the maximizing coupling is comonotone and the minimizing coupling is countermonotone. Substituting the diagonal and anti-diagonal weights into (16) yields the stated bounds. □

A.4 Contraction mapping under serial-copula shocks

This appendix records a standard contraction argument for the dynamic program when serial dependence in utility shocks is represented by the copula-induced latent Markov state.

Assumption CM (Discounting, bounded flow utility, and Markov state). *Let the augmented state be $s \equiv (x,k) \in \mathcal{S}$, where $x \in \mathcal{X}$ is the observed state and $k \in \mathcal{K}$ is a latent dependence index. Assume: (i) $\beta \in (0,1)$; (ii) $\sup_{x\in\mathcal{X}, j\in\{1,\ldots,J\}} |u(x,j)| < \infty$; (iii) conditional on (s,j), the next augmented state $s' = (x',k')$ follows a Markov transition kernel $P(\cdot \mid s,j)$.*

Under Assumption CM, define the ex-ante value function on $\mathcal{S}$ by

$$G(s) \equiv \mathbb{E}\Big[\max_{j\in\{1,\ldots,J\}} \{u(x,j) + \varepsilon_j + \beta\,\mathbb{E}[G(s') \mid s,j]\}\Big|\, s\Big],$$

where the inner expectation is taken with respect to $P(\cdot \mid s,j)$ and the outer expectation integrates over the within-period shock vector ε conditional on s.

Proposition 3 (Contraction on the augmented state space). *Under Assumption CM, the Bellman operator*

$$(TG)(s) \equiv \mathbb{E}\Big[\max_{j\in\{1,\ldots,J\}} \{u(x,j) + \varepsilon_j + \beta\,\mathbb{E}[G(s') \mid s,j]\}\Big|\, s\Big]$$

is a contraction mapping on the space of bounded functions on $\mathcal{S}$ endowed with the sup norm, with modulus β. Hence there exists a unique fixed point $G^\star$ satisfying $G^\star = TG^\star$, and value function iteration $G^{(n+1)} \equiv TG^{(n)}$ converges to $G^\star$ at geometric rate β.

Proof. Let G_1, G_2 be bounded functions on $\mathcal{S}$. For each $s = (x,k)$ and each realisation of ε,

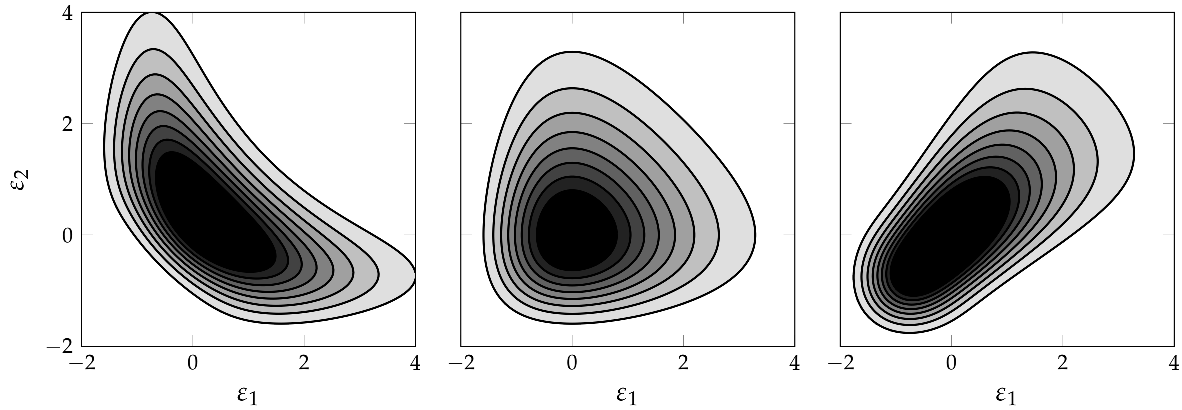
$$\Big|\max_j\{u(x,j) + \varepsilon_j + \beta\mathbb{E}[G_1(s') \mid s,j]\} - \max_j\{u(x,j) + \varepsilon_j + \beta\mathbb{E}[G_2(s') \mid s,j]\}\Big| \\ \le \beta \max_j \Big|\mathbb{E}[G_1(s') - G_2(s') \mid s,j]\Big|.$$

By Jensen's inequality and boundedness, $\big|\mathbb{E}[G_1(s') - G_2(s') \mid s,j]\big| \le \mathbb{E}[|G_1(s') - G_2(s')| \mid s,j] \le \|G_1 - G_2\|_\infty$. Taking the maximum over j and then taking expectations over ε conditional on s yields $|(TG_1)(s) - (TG_2)(s)| \le \beta\,\|G_1 - G_2\|_\infty$ for all s, and hence $\|TG_1 - TG_2\|_\infty \le \beta\|G_1 - G_2\|_\infty$. The fixed point and convergence claims follow. □

Remark. Proposition 3 is a generic dynamic-programming result, with the contraction property following from discounting and a Markov state. The serial-copula specification in Assumption BC provides a finite latent index K_t with an explicit Markov transition matrix implied by the copula parameters. This keeps the dynamic program on the finite augmented state space (x_t, K_t), so the contraction fixed point can be computed with standard methods.

B Dependence properties under low-dimensional weights

In Section 3.6 we introduced a convenient one-parameter weight specification for the bivariate Bernstein copula. Figure 1 plots contour maps of the implied joint density of $(\varepsilon_1, \varepsilon_2)$ at degree $m = 5$ for $\theta \in \{-1, 0, 1\}$ (negative dependence, independence, and positive dependence; see Corollary 1). The main visual takeaway is that θ rotates the high-density region: for $\theta > 0$ the contours elongate along the diagonal (persistence), while for $\theta < 0$ they elongate along the



(a) One param ($\rho = -1$) (b) One param ($\rho = 0$) (c) One param ($\rho = 1$)

Figure 1: Bernstein copula with one-parameter weight specification ($m = 5$). Panels (a)–(c) correspond to $\theta \in \{-1, 0, 1\}$ in (17). Contours are density level sets drawn at reference thresholds chosen so that, under independence (panel (b)), each contour encloses probability $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$. The same thresholds are reused in panels (a) and (c) to make the change in contour geometry directly comparable across dependence levels. Higher α corresponds to a lower density threshold and hence a larger enclosed region.

anti-diagonal (mean reversion), corresponding to latent-rank transitions concentrating near $K_2 \approx K_1$ versus $K_2 \approx m + 1 - K_1$.

In Figure 2 we illustrate the quality of the Bernstein approximation to a target copula using contour maps of the joint density. The target is a Gaussian copula with correlation coefficient 0.6 (rank correlation $\rho_S \approx 0.58$). Panels (a) and (b) show Bernstein approximations at degrees $m = 2$ and $m = 8$ where the weights have been calculated using (19), while panel (c) shows the target. Increasing m improves the match of contour geometry to the target, reflecting a finer latent-rank discretisation.

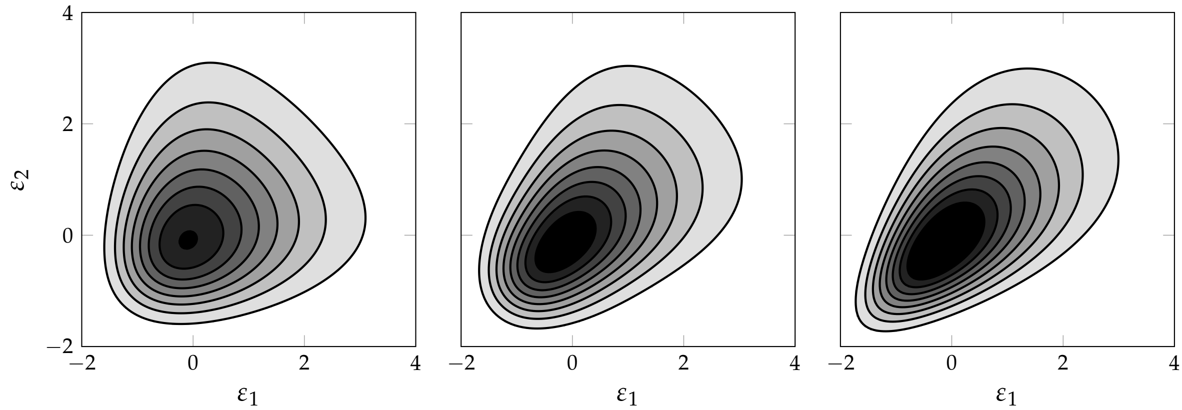
C Across-alternative (within-period) dependence

The main text focuses on serial dependence within an alternative over time. The same framework can also accommodate across-alternative (within-period) dependence among shocks within a period. We now describe this extension in a static setting for simplicity.

C.1 A Bernstein copula for across-alternative dependence

Consider a multinomial choice model where the decision-maker chooses among I alternatives with utilities $v_j + \varepsilon_j$ for $j = 1, \ldots, I$, where $\{v_j\}$ are deterministic value indices and $\{\varepsilon_j\}$ are random shocks with T1EV marginals. Across-alternative (within-period) dependence among the shocks can influence substitution patterns across alternatives and can be modeled using copulas. Consider a Bernstein copula.

Assumption 1 (Across-alternative Bernstein copula)**.** *Fix an integer degree $m \geq 1$ and let $B_{r,m}(u)$*



(a) Bernstein copula ($m = 2$) (b) Bernstein copula ($m = 8$) (c) Gaussian copula ($\rho = 0.6$)

Figure 2: Bernstein approximation to a Gaussian copula. Panels (a) and (b) show Bernstein approximations based on (19) at degrees $m = 2$ and $m = 8$, respectively; panel (c) shows the Gaussian target (correlation 0.6). Contours are density level sets drawn at reference thresholds chosen so that, under the target (panel (c)), each contour encloses probability $\alpha \in \{0.1, 0.2, \ldots, 0.9\}$. The same thresholds are reused in panels (a) and (b); closer approximation corresponds to closer alignment of contour geometry with the target.

denote the Beta CDF for $r = 1, \ldots, m$. The copula for $\mathbf{U} = (U_1, \ldots, U_I)$ is the Bernstein mixture

$$C(u_1, \ldots, u_I) = \sum_{k_1=1}^{m} \cdots \sum_{k_I=1}^{m} w_{k_1,\ldots,k_I} \prod_{j=1}^{I} B_{k_j,m}(u_j), \tag{C.1}$$

where $W \equiv (w_{k_1,\ldots,k_I})$ satisfies

$$w_{k_1,\ldots,k_I} \geq 0, \qquad \sum_{k_{-j}} w_{k_1,\ldots,k_I} = \frac{1}{m} \text{ for each } j \text{ and each } k_j \in \{1, \ldots, m\}, \tag{C.2}$$

where $\sum_{k_{-j}}$ denotes summation over all indices except k_j.

The constraints (C.2) are linear. They enforce that each K_j is uniform on $\{1, \ldots, m\}$ (hence each U_j is marginally uniform), while allowing arbitrary dependence across j through the joint weight tensor $(w_{k_1,\ldots,k_I})$.

The formal structure is analogous to that in the serial dependence case, and we state the following Proposition without proof.

Proposition 4 (CCPs and inclusive value under the across-alternative Bernstein copula)**.** *Fix degree $m \geq 1$ and let $\boldsymbol{v} \equiv (v_1, \ldots, v_I)$ denote the vector of choice-specific value indices. Then the CCP is the finite mixture*

$$\mathbb{P}\{y = j \mid \boldsymbol{v}\} = \sum_{k_1=1}^{m} \cdots \sum_{k_I=1}^{m} w_{k_1,\ldots,k_I} \sum_{q_1=0}^{m-k_1} \cdots \sum_{q_I=0}^{m-k_I} \left[\prod_{i=1}^{I} \kappa_{k_i,q_i}\right] \frac{\lambda_{k_j,q_j} e^{v_j}}{\sum_{\ell=1}^{I} \lambda_{k_\ell,q_\ell} e^{v_\ell}}, \tag{C.3}$$

where $\{\lambda_{sq}\}$ *and* $\{\kappa_{sq}\}$ *are defined in* (5) *and the inclusive value is*

$$G(\boldsymbol{v}) = \gamma + \sum_{k_1=1}^{m} \cdots \sum_{k_I=1}^{m} w_{k_1,\dots,k_I} \sum_{q_1=0}^{m-k_1} \cdots \sum_{q_I=0}^{m-k_I} \left[\prod_{i=1}^{I} \kappa_{k_i,q_i} \right] \log \left(\sum_{\ell=1}^{I} \lambda_{k_\ell,q_\ell} e^{v_\ell} \right). \tag{C.4}$$

Conditional on the latent rank vector K, the shocks admit a finite expansion (after the transformation $Z_j = -\log U_j$), and each expansion term corresponds to an independent exponential race with rates λ_{K_j,q_j}. Thus, conditional on (K, q) the CCPs are rate-shifted logits and the inclusive value is a corresponding log-sum with rate shifts. Across-alternative dependence enters only through the weights $w_{k_1,\dots,k_I}$.

In the serial dependence model, we assumed within-period independence across alternatives and encoded dependence only over time (Assumption SC). This is computationally convenient because it avoids an m^I-dimensional within-period weight tensor. In contrast, across-alternative dependence requires a full m^I-dimensional weight tensor $(w_{k_1,\dots,k_I})$. Evaluating (C.3) and (C.4) then involves sums over $K \in \{1, \dots, m\}^I$ (size m^I) and, for each K, sums over multi-indices $q \in \prod_{j=1}^{I} \{0, \dots, m - k_j\}$. The operation count is therefore approximately $O((m^2/2)^I)$, which becomes computationally challenging even for moderate m and I. Structured restrictions (e.g. block/nest structure, or sparse pairwise interactions) can help mitigate this curse of dimensionality.

Example (red bus / blue bus). Consider the classic "red bus / blue bus" thought experiment (McFadden, 1974). Initially, there are two transport modes, car (C) and a red bus (R). Deterministic utilities are $v_C = v_R = 0$, and the shocks are i.i.d. Type-I extreme value. Then $p_C = p_R = 1/2$. Now introduce a blue bus (B) that is identical to the red bus apart from colour, also with $v_B = 0$. Under i.i.d. extreme-value errors, IIA implies $p_C = p_R = p_B = 1/3$, which is the familiar counterfactual substitution pattern.

While a nested logit model is the most natural way to relax IIA here, for illustrative purposes we show how the Bernstein construction can also achieve this. Let $U_j \equiv F(\varepsilon_j)$ be the copula-scale variables, and take $U_C \sim \mathrm{Unif}(0, 1)$ independent of (U_R, U_B). Similar to (17), we use a one-parameter family to structure dependence for the bus pair

$$w^{(RB)}(\theta) = (1 - \theta) w^{\mathrm{ind}} + \theta w^{+}, \qquad \theta \in [0, 1],$$

where $w^{\mathrm{ind}}_{rs} = m^{-2}$ (independence weights) and $w^{+}_{rs} = m^{-1}\, \mathbb{1}[r = s]$ (diagonal weights). Obviously this can be extended to accommodate negative dependence, but that is not our focus here. Because the Bernstein CCP expression (C.3) is linear in the weights, these market shares are affine in θ.[20] Table C.1 illustrates how the market shares vary with θ for alternative values of m. When $\theta = 0$ we recover i.i.d. errors and hence $p_C = p_R = p_B = 1/3$. As θ increases, the two buses share a more common latent rank index and become closer substitutes, so the car share rises. In the limit $\theta \to 1$ and m large, (U_R, U_B) becomes close to comonotone and the

[20] In generative terms, first draw $K_R \sim \mathrm{Unif}\{1, \dots, m\}$. Then draw $K_B = K_R$ with probability θ (the diagonal component w^+), and draw $K_B \sim \mathrm{Unif}\{1, \dots, m\}$ with probability $1 - \theta$ (the independence component w^{ind}). Next draw $U_R \mid K_R \sim \mathrm{Beta}(K_R, m - K_R + 1)$ and $U_B \mid K_B \sim \mathrm{Beta}(K_B, m - K_B + 1)$, independently conditional on (K_R, K_B), and set $\varepsilon_j = -\log(-\log U_j)$.

	$m=4$		$m=16$		$m=32$	
θ	p_C	$p_R = p_B$	p_C	$p_R = p_B$	p_C	$p_R = p_B$
0	0.333	0.333	0.333	0.333	0.333	0.333
0.5	0.366	0.317	0.390	0.305	0.397	0.301
1.0	0.398	0.301	0.446	0.277	0.461	0.269

Table C.1: Market shares in the red bus / blue bus example under the across-alternative Bernstein copula with dependence parameter $\theta \in [0,1]$. Columns report p_C and $p_R = p_B$ for different degrees m at $\theta \in \{0, 0.5, 1\}$.

car share approaches the original two-alternative value $1/2$ (with the two buses splitting the remaining half).

C.2 Discussion: alternative copulas and tradeoffs

Alternative copula classes are also convenient to model across-alternative (within-period) dependence. In contrast to serial dependence models, a latent index structure is not required for tractability. Many existing discrete-choice models with within-period dependence can be reinterpreted in copula terms. For example, McFadden's GEV class (McFadden, 1978, 1981) corresponds to multivariate extreme-value shock distributions, and therefore to extreme-value copulas. One such example is the nested logit model, which has convenient closed-form expressions, but restricts dependence to specific patterns tied to the GEV structure.

A simple polynomial copula class is provided by the *Farlie–Gumbel–Morgenstern* (FGM) family (Morgenstern, 1956; Gumbel, 1958; Farlie, 1960). The polynomial structure leads, under T1EV marginals, to CCP and inclusive-value expressions that remain in closed form.[21] Indeed, it is possible to express the resulting closed-form objects as finite linear combinations of logit choice probabilities and log-sum expressions that arise under independence under an augmented choice set in which some alternatives are effectively duplicated. Validity can be enforced using finitely many linear inequalities in the dependence parameters (Johnson and Kotz, 1975). The main drawback is that attainable dependence is weak (Schucany et al., 1978) and it becomes increasingly restrictive as the number of alternatives grows.[22]

More generally, copulas with polynomial densities lead, under T1EV marginals, to CCP and inclusive-value expressions that remain in closed form because the relevant integrals reduce to a finite set of logit-like terms. In many such polynomial specifications (for example, a copula density formed by using a shifted Legendre basis), higher-degree terms correspond to allowing more "copies" of alternatives in the augmented-choice-set representation, so increasing polynomial order enriches the dependence patterns that can be represented while retaining finite closed-form evaluation. The tradeoff is feasibility. Away from special parameterizations, ensuring that a polynomial density remains nonnegative over $[0,1]^I$ typically

[21] More generally, because the FGM density is polynomial in the marginal CDFs, any marginal specification that yields closed-form CCPs and inclusive values under independence inherits closed forms under FGM. This includes logistic marginals and semi-nonparametric Gumbel-based density expansions built from orthonormal Legendre polynomials (Bierens, 2008).

[22] In the context of the red bus / blue bus example discussed above, at the maximal dependence level between the red and blue buses, the FGM copula delivers $p_C = 22/60 \approx 0.367$ and $p_R = p_B = 19/60 \approx 0.317$.

imposes constraints that are considerably more complex than the Bernstein weight restrictions (see Nelsen, 2006, for background on polynomial copulas).

Finally, many popular copula choices (e.g., Gaussian copulas) typically lead to CCPs and inclusive values without closed forms, so they require numerical integration or simulation.

D Identification

This Appendix provides a more formal identification argument for the serial copula model. The key observation is that, conditional on observed states, the model induces a finite-state hidden Markov model for observed choices: the latent state is the vector of copula rank indices $K_t \in \{1, \ldots, m\}^I$, the latent transition matrix is the copula-implied $\Pi(\theta)$, and the emission probabilities are the structural conditional choice probabilities derived in Proposition 1. The argument proceeds in three steps:

1. Identify the reduced-form HMM objects (transition and emissions) from the joint distribution of observables using existing HMM identification results;
2. Resolve the generic label-switching indeterminacy by exploiting the fact that our latent states have a canonical economic interpretation as ordered rank regions;
3. Map the labelled reduced-form objects back to the structural primitives (utility parameters and copula dependence parameter).

Setup. Let $y_t \in \{1, \ldots, I\}$ denote the period-t discrete choice and let x_t denote the observed state vector. Assumptions AS, CL, SC, BC, and CI are maintained. That is, the additively separable unobservables are Type-I extreme value utility shocks and their serial dependence can be represented using the Bernstein-copula-induced rank state $K_t \in \{1, \ldots, m\}^I$.

To focus on identification, it is helpful to separate (i) the reduced-form objects governing the joint law of (y_t, K_t) conditional on x_t, from (ii) the structural restrictions imposed by the copula and the dynamic program.

Assumption 2 (Short-panel HMM structure)**.** *Conditional on x_t, the process (y_t, K_t) is first-order Markov with*

$$\begin{aligned} \mathbb{P}\{y_t = j \mid x_t, K_t = k, \mathcal{F}_{t-1}\} &= \mathbb{P}\{y_t = j \mid x_t, K_t = k\}, \\ \mathbb{P}\{K_{t+1} = k' \mid x_{t+1}, x_t, K_t = k, \mathcal{F}_{t-1}\} &= \mathbb{P}\{K_{t+1} = k' \mid K_t = k\}, \end{aligned}$$

where $\mathcal{F}_{t-1}$ collects all information dated $t-1$ and earlier.

Assumption 2 is the "limited feedback" structure used in the hidden Markov and hidden-Rust literatures: conditional on the current latent state, past latent states affect current outcomes only through the current one. In our model this is built in: conditioning on K_t restores within-period tractability and the copula construction implies a Markov transition for K_t.[23]

[23]The formal argument in this Appendix is written for first-order serial dependence, where the latent rank state is $K_t \in \{1, \ldots, m\}^I$. The higher-order serial dependence extension discussed in Section 3.7 does not change the logic: an Lth-order process can always be rewritten as a first-order Markov process by augmenting the latent state to $S_t \equiv (K_{t-L+1}, \ldots, K_t)$. One then applies the same three identification steps with S_t in place of K_t.

Define the reduced-form HMM objects:

$$B_x(j \mid k) \equiv \mathbb{P}\{y_t = j \mid x_t = x, K_t = k\}, \qquad \Pi(k,k') \equiv \mathbb{P}\{K_{t+1} = k' \mid K_t = k\},$$

and an initial distribution $\alpha_1(k) \equiv \mathbb{P}\{K_1 = k\}$.

Step 1: reduced-form identification of (Π, B_x)

Identification of (Π, B_x) from the joint distribution of $\{(y_t, x_t)\}_{t=1}^T$ is a standard question in the HMM literature. See, for example, Allman et al. (2009), Hsu et al. (2012), Hu and Shum (2012), and Connault (2016). For our purposes we do not need to restate the most general versions of these results. What we use is the implication that, under standard rank/distinctness conditions, the reduced-form HMM is identified from a short panel up to a permutation of the latent-state labels.

Proposition 5 (Reduced-form identification up to permutation)**.** *Suppose Assumption 2 holds and the latent state space $\mathcal{K} \equiv \{1,\ldots,m\}^I$ is finite. Assume further that (i) the latent chain is ergodic and its transition matrix Π has full rank, and (ii) for at least one value x in the support of x_t the emission matrix $\big(B_x(j \mid k)\big)_{j\in\{1,\ldots,I\},k\in\mathcal{K}}$ satisfies a full-rank/distinctness condition (as in Allman et al., 2009 and Hsu et al., 2012).*

Then the reduced-form objects $\big(\alpha_1, \Pi, \{B_x\}_{x\in\mathrm{supp}(x_t)}\big)$ are identified from the joint distribution of observable histories $\{(y_t, x_t)\}_{t=1}^T$ for a finite T ("short panels"), up to a common permutation of the latent-state labels.

Proposition 5 captures the econometric content of the first step: from observed choice histories (and conditioning on observed states), one can recover the latent transition and the state-dependent CCPs, but only up to relabelling of the latent states. This is the familiar label-switching indeterminacy in mixture and HMM models.

Step 2: resolving label switching using rank-state structure

In an unrestricted HMM, the latent labels have no inherent meaning and the permutation ambiguity is fundamental. In the serial copula model, however, each latent state $k \in \mathcal{K}$ corresponds to an *ordered* rank region of the copula-transformed shocks. This structure provides economically meaningful restrictions that can be used to pin down the permutation.

To formalize this, write $k = (k_1, \ldots, k_I)$ with $k_j \in \{1,\ldots,m\}$, and define the partial order $k \preceq k'$ if $k_j \le k'_j$ for all j.

Assumption 3 (Monotone rank labelling)**.** *For each alternative j and each observed state value x, the conditional choice probability $B_x(j \mid k)$ is weakly increasing in k_j and weakly decreasing in k_ℓ for $\ell \neq j$, with at least one strict inequality on a set of (x,k) values of positive probability.*

Assumption 3 is a reduced-form implication of the underlying construction: higher rank index k_j corresponds to a stochastically larger shock for alternative j (holding the other coordinates fixed), so alternative j becomes more likely to be chosen.

Proposition 6 (Canonical ordering of latent states)**.** *Under the conditions of Proposition 5 and Assumption 3, the permutation indeterminacy can be resolved: there exists a unique relabelling of the identified latent states such that the monotonicity restrictions in Assumption 3 hold.*

In words, the copula rank interpretation provides a canonical coordinate system for the latent state space, so that the reduced-form HMM can be anchored to the structural latent ranks. This is the sense in which the copula representation delivers more than a generic latent-regime model.

Step 3: mapping (Π, B_x) to structural primitives

Once the reduced-form objects are labelled, identification of economic primitives proceeds by exploiting the structural restrictions imposed by the copula and the dynamic program.

(i) Identifying the dependence parameter θ. The copula specification restricts the latent transition matrix to lie in a low-dimensional family $\{\Pi(\theta) : \theta \in \Theta\}$.

Assumption 4 (Injective transition parameterization)**.** *The map $\theta \mapsto \Pi(\theta)$ is injective on Θ.*

Under Assumption 4, the labelled transition matrix Π that was recovered in Step 2 point-identifies θ.

(ii) Identifying flow-utility parameters from labelled emissions. Conditional on (x_t, K_t), our model implies that the CCPs are given by the closed-form mixture-logit expression in (7), where the indices $\tilde{v}_{jt}$ are measurable functions of (x_t, K_t) determined by the structural flow utility and continuation value.

From an identification standpoint, the key point is that conditional on $K_t = k$ the model is a standard dynamic discrete choice problem on an *augmented* observed state (x_t, k). Therefore, any identification strategy for i.i.d.-logit dynamic discrete choice models applies conditional on k. In particular, CCP-based identification arguments (e.g. Hotz and Miller, 1993; Magnac and Thesmar, 2002) imply that, given the discount factor β and the law of motion for x_{t+1}, the choice-specific value indices are identified (up to standard normalizations) from the family of conditional CCPs $\{B_x(\cdot \mid k)\}$, and hence so are the underlying payoff parameters entering the flow utility function $u(x, j)$.

Remark. As in the classic i.i.d.-logit case, normalizations are required. In our setting the T1EV marginal normalization fixes the shock scale, so remaining normalizations are the usual location/level normalizations of deterministic utilities. If the discount factor β is treated as unknown, additional restrictions are required, exactly as in standard dynamic discrete choice models (see Magnac and Thesmar, 2002).

E Additional results

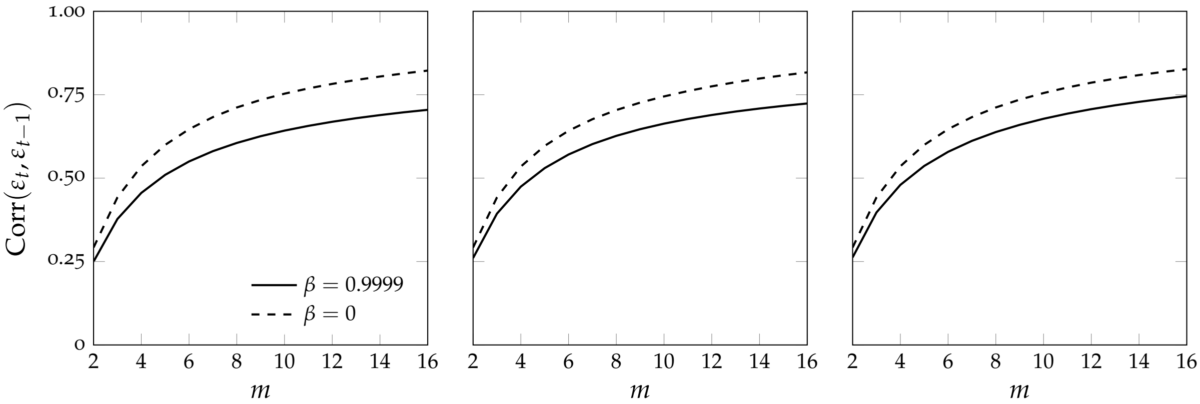
In Section 4 we presented Monte Carlo results for the serial copula model estimated under correct specification. In Table E.1 we report complementary results showing the performance

n	T	Positive dep. ($\theta = 0.8$)		Independence ($\theta = 0.0$)		Negative dep. ($\theta = -0.8$)	
		$\hat{\omega}_2$	$\hat{h}$	$\hat{\omega}_2$	$\hat{h}$	$\hat{\omega}_2$	$\hat{h}$
100	3	0.10	1.03	0.20	0.40	0.28	0.22
		(0.08)	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)
100	5	0.08	0.71	0.20	0.40	0.29	0.37
		(0.05)	(0.05)	(0.07)	(0.04)	(0.07)	(0.05)
100	10	0.06	0.44	0.20	0.40	0.26	0.40
		(0.03)	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)
1000	3	0.11	1.03	0.20	0.40	0.28	0.23
		(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
1000	5	0.08	0.71	0.20	0.40	0.28	0.37
		(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)
1000	10	0.06	0.43	0.20	0.40	0.26	0.39
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Table E.1: Monte Carlo results under misspecified i.i.d.-logit estimation (θ restricted to 0). Data generated with parameters $(\omega_2, h) = (0.2, 0.4)$ and varying θ. Each cell reports the mean estimate and average standard error over 1000 simulations. Here n is the cross-sectional sample size and T is the panel length.

of a misspecified i.i.d.-logit estimator that constrains the dependence parameter to $\theta = 0$. The data are generated from the same DGP as in Section 4, with parameters $(\omega_2, h) = (0.2, 0.4)$ and varying θ. The table reports the mean estimate and average standard error over 1000 simulations for each (n, T, θ) combination.

Section 5 presented a version of the Rust (1987) bus engine replacement model with serial dependence. In Figure 3 we present the correlation coefficient associated with the estimated dependence parameter (the correlation coefficient of the target Gaussian copula, as described in Section 3.6) as the copula degree m is varied. The different panels correspond to the different bus group samples, and within each panel we show results for both $\beta = 0.9999$ (solid line) and $\beta = 0$ (broken line). In all cases, the estimated correlation coefficient increases with m, reflecting the greater dependence that can be captured with higher-degree Bernstein copulas, and is higher with $\beta = 0$. The initial increase is particularly steep, with estimates stabilizing from around $m \approx 6$.



(a) Groups 1–3

(b) Group 4

(c) Groups 1–4

Figure 3: Implied serial dependence under the Bernstein-copula specification: $\mathrm{Corr}(\varepsilon_t, \varepsilon_{t-1})$ vs copula degree m for the Rust (1987) bus engine replacement model (Groups 1–3; Group 4; Groups 1–4). Solid (dashed) line shows values with $\beta = 0.9999$ ($\beta = 0$).