

Dynamic Panel Multinomial Logit Models ^{*}

Kevin Dano[‡]

Bo E. Honoré[§]

Martin Weidner[¶]

October 7, 2026

Abstract

This paper provides moment conditions for dynamic multinomial logit panel data models with individual specific fixed effects. The moment conditions do not depend on the fixed effects and thus form the basis for generalized method of moments estimation. For a leading case, we provide sufficient conditions under which the moment conditions identify the common parameters, and we illustrate their use in an empirical application to household soft drink purchases. The construction of moment conditions is based on transformed models obtained by permuting outcomes across time periods. For any cyclic permutation, the ratio of the transformed model probability to the true model probability is free of the fixed effects, and comparing two such ratios yields restrictions on the common parameters of interest. This construction is not specific to the multinomial logit. It applies to any dynamic panel data model whose static counterpart admits a symmetric sufficient statistic for the fixed effects. This unifies our main result with analogous results for other models.

^{*}This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University and by the Danish National Research Foundation grant number DNR186 to the Aarhus Center for Econometrics (ACE). We are grateful to seminar participants at numerous institutions for constructive comments.

[‡]Princeton University, kdano@princeton.edu

[§]Princeton University and Aarhus Center for Econometrics (ACE) at the University of Aarhus, honore@princeton.edu

[¶]University College London, m.weidner@ucl.ac.uk

1 Introduction

Controlling for individual heterogeneity without imposing strong assumptions on its relationship to observed covariates is a central challenge in the econometric analysis of nonlinear panel data models. Suppose the individual-specific effects are treated as parameters and estimated jointly with the common parameters of interest. The resulting estimator is then generally inconsistent as the number of cross-sectional units grows with the number of time periods held fixed. This is the classical incidental parameter problem ([Neyman and Scott, 1948](#)). This paper studies the dynamic multinomial logit model for panel data with fixed effects. Our main contribution is to construct moment conditions that depend on the common parameters but not on the individual-specific effects. We obtain them from a transformed model that permutes outcomes across time periods, and that cancels the fixed effects through the structure of the multinomial logit likelihood.

One approach to panel data discrete choice models with fixed effects is to look for sufficient statistics for the fixed effects. It dates back to [Rasch \(1960\)](#), who observed that in a binary logit model with strictly exogenous regressors the sum of an individual's outcomes is a sufficient statistic for the fixed effect, so that conditional maximum likelihood can be used. [Chamberlain \(1980\)](#) extended this to the static multinomial logit, where the vector of choice frequencies is sufficient for the fixed effects, and [Magnac \(2000\)](#) obtained an analogous result when the only regressors are lagged dependent variables. Unfortunately, conditional likelihood methods generally break down in the empirically most relevant case in which both lagged dependent variables and strictly exogenous regressors are present. [Honoré and Kyriazidou \(2000\)](#) handled this joint case for binary and multinomial logit by additionally conditioning on covariates being equal across certain periods. This leads to an estimator whose rate of convergence is slower than $\sqrt{n}$. [Kyriazidou \(2001\)](#) combined conditional likelihood with kernel smoothing in dynamic panel sample selection models. [Khan, Ouyang and Tamer \(2021\)](#) generalized [Honoré and Kyriazidou \(2000\)](#) to semiparametric multinomial response models that drop the logit assumption and relax the independence of irrelevant alternatives, establishing point identification but again at a rate slower than $\sqrt{n}$ that deteriorates with the number of alternatives.

Since there are relatively few models for which sufficient statistics can be found, an alternative approach is to try to construct moment conditions that depend on the common parameters, but not on the individual-specific fixed effects. In the linear case this is what [Arellano and Bond \(1991\)](#) do, when they eliminate the fixed effects by first-differencing and instrument with lagged levels. Papers by [Honoré \(1993\)](#) and [Hu \(2002\)](#) are specific examples of this for panel Tobit models, where the conditions are obtained from a symmetry that is restored through trimming, and the argument is tailored to the censored-regression structure rather than being generic. For multiplicative models, [Chamberlain \(1992\)](#) and [Wooldridge \(1997\)](#) derived moment conditions which are valid without strict exogeneity, implying that lagged dependent variables can be accommodated in count and other nonnegative outcome models. [Bonhomme \(2012\)](#) developed a general approach for obtaining moment conditions via functional differencing, which essentially differences out the unknown distribution of the individual effects in the likelihood. While general, this characterization does not by itself deliver explicit moment functions, and concrete constructions have so far proceeded model by model, exploiting the specific structure of each likelihood. Functional differencing has been operationalized for the dynamic binary logit model by [Kitazawa \(2022\)](#) and [Honoré and Weidner \(2025\)](#), extended to the dynamic ordered logit by [Honoré, Muris and Weidner \(2025\)](#), and given a general transition-probability treatment by [Dano \(2023\)](#). The present paper extends this program to the dynamic multinomial logit.¹

Multinomial outcomes are common in applied work with panel data. They arise in choices among employment states ([Keane and Wolpin, 1997](#)), consumer products with repeat purchasing and switching costs ([Handel, 2013](#); [Dubé, Hitsch and Rossi, 2010](#)), health insurance plans ([Pakes, Porter, Shepard and Calder-Wang, 2021](#)), urban consumption locations ([Davis, Dingel, Monras and Morales, 2019](#)), and migration destinations ([Blumenstock, Chi and Tan, 2025](#)). All of these fit naturally into the multinomial framework. Because switching costs and habit formation are prominent in such settings, it is important to allow lagged choices to enter as regressors and to control for permanent unobserved heterogeneity at the same time

¹[Al-Sadoon, Li, Pesaran, Yang and Zhao \(2026\)](#) also consider moment conditions in a dynamic multinomial fixed effects model, but their model is outside the multinomial logit framework and it places restrictions on the support of the fixed effects that depend on the explanatory variables and the model parameters.

([Heckman, 1981](#)).

The construction of our moment conditions is based on a transformed model in which the outcome generated in each period is reassigned according to a permutation σ of the time indices. For any σ , the ratio of the transformed-model probability to the true-model probability does not depend on the fixed effects. This follows from the logistic structure, in which the fixed effects enter the choice probabilities via alternative-specific constants whose total contribution depends on the unordered collection of outcomes only — implying invariance to reshuffling outcomes across time. Intuitively, reshuffling outcomes across time leaves the total contribution of the individual effects unchanged, so they can be differenced out through likelihood ratios. When σ is a cyclic permutation the transformed model is a proper probability distribution, and comparing two such models yields moment conditions of the form required for fixed effects-free estimation; the binary logit conditions of [Honoré and Weidner \(2025\)](#) emerge as the special case with two alternatives. The construction is not specific to the multinomial logit. All that is needed is that the static version of the model admits a sufficient statistic for the fixed effects that is symmetric in the time indices, i.e. a statistic that depends on the outcomes only through their unordered collection. This unifies our results with analogous ones for other models and shows that the approach applies to any dynamic panel model built on a static model that admits conditional likelihood estimation via a symmetric sufficient statistic.

The number of linearly independent moment conditions for given (Q, T) grows rapidly in both the number of alternatives Q and the number of time periods T , and we conjecture an explicit formula for this count. We prove that this formula is correct for $Q = 2$, and that it is a lower bound for $T = 3$ and any Q . For the other combinations of (Q, T) that we report we verify it numerically. At least three time periods after the initial condition are required, as in the binary case. We give sufficient conditions under which these moments identify the common parameters, the regression and state-dependence coefficients, and propose a generalized method of moments (GMM) estimator. Because the number of available moments grows quickly, only a small subset is typically used in estimation. An empirical application to household soft drink purchases illustrates the importance of controlling for unobserved

heterogeneity when estimating state dependence in multinomial choice.

The paper is organized as follows. Section 2 introduces the model. Section 3 presents the moment conditions and establishes their validity. Section 4 generalizes the construction to other dynamic models whose static versions admit a symmetric sufficient statistic for the fixed effects. Section 5 analyzes identification. Section 6 contains the empirical application. Section 7 concludes.

2 Multinomial logit model

Panel settings in which agents repeatedly select from a finite set of alternatives motivate our framework. Examples of such alternatives are brands of a consumer good, transportation modes, employment states, or health insurance plans. Our analysis focuses on two features of these settings: agents differ in their persistent preferences over alternatives (a variation not observed by the econometrician), and past choices can influence current utility (which gives rise to state dependence). Both features produce persistence in the observed choices, so distinguishing them is difficult on the basis of cross-sectional data alone. This is why we use panel data where the same agents are observed over multiple time periods.

We consider panel data with cross-sectional units $i = 1, \dots, n$ and time periods $t = 0, \dots, T$. For each pair (i, t) , we observe the discrete outcome $Y_{it} \in \{1, 2, \dots, Q\}$, where $Q \in \{2, 3, 4, \dots\}$ denotes the number of options. We also observe strictly exogenous regressors $X_{it} \in \mathbb{R}^K$ for $t \geq 1$. The observed outcome at $t = 0$ serves as the initial condition and is conditioned on throughout. For $t \geq 1$, the outcomes are generated by the model

$$Y_{it} = \underset{q \in \{1, 2, \dots, Q\}}{\operatorname{argmax}} U_{qit}, \quad (1)$$

with latent utilities

$$U_{qit} = z_t(q, Y_{i,t-1}, X_i, \theta) + A_{qi} + \varepsilon_{qit}. \quad (2)$$

Here, $A_i = (A_{1i}, \dots, A_{Qi})$ are individual-specific fixed effects, and $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})$

are all the strictly exogenous covariates observed for unit i . The errors ε_{qit} are independent of (Y_{i0}, X_i, A_i) , and are assumed to be independent and identically distributed (i.i.d.) over t and q with a type I extreme value distribution.

The vector A_i collects alternative-specific constants for unit i , with A_{qi} capturing the taste for option q . Since only differences between the A_{qi} enter the choice probabilities, one can normalize $A_{1i} = 0$. The type I extreme value assumption imply that the choice probabilities in (5) below take the form of the familiar multinomial logit model. The independence of irrelevant alternatives across alternatives within each period is therefore the key substantive restriction. Conditioning on Y_{i0} throughout avoids the need to specify a model for the initial outcome. This is convenient, because Y_{i0} is typically correlated with the fixed effects in an unknown way.

The parameter θ is unknown, but the function $z_t(\cdot, \cdot, \cdot, \theta)$ is known up to θ . The standard linear specification reads

$$z_t(q, Y_{i,t-1}, X_i, \theta) = \sum_{r=1}^Q \gamma_{qr} \mathbb{1}\{Y_{i,t-1} = r\} + X'_{it} \beta_q, \quad (3)$$

with $\theta = (\gamma, \beta)$. Our primary interest is in the state dependence coefficients γ_{qr} and the regression coefficients β_q . The diagonal entries γ_{qq} measure how much the utility of alternative q rises when q was also chosen in the previous period. In other words, γ_{qq} captures the degree of habit formation or switching costs for option q . The off-diagonal entries γ_{qr} with $r \neq q$ capture additional forms of persistence, including asymmetries between related alternatives. The central difficulty in estimating the persistence coefficients is that one would like to separate genuine state dependence (in the form of the γ 's) from spurious persistence induced by unobserved heterogeneity. See, for example, [Heckman \(1981\)](#). We address this problem by constructing moment conditions that are free of A_i but informative about the γ 's.

The coefficients in the model above are identified only up to normalizations. Three types of shifts leave the choice probabilities unchanged: (i) shifting γ_{qr} by a constant that depends on r but not on q , because such a shift appears in the numerator and in every term of the denominator of the logit and therefore cancels; (ii) shifting γ_{qr} by a constant that depends

on q but not on r , because $\sum_{r=1}^Q \mathbb{1}\{Y_{i,t-1} = r\} = 1$ and such a shift is therefore equivalent to adding a constant to the utility of alternative q , which is absorbed into the fixed effect A_{qi} ; and (iii) shifting β_q by an additive vector in $\mathbb{R}^K$ that does not depend on q , which cancels between the numerator and denominator of the logit. To remove these sources of non-identification we impose the normalizations $\gamma_{1r} = 0$ for all $r \in \{1, \dots, Q\}$, $\gamma_{q1} = 0$ for all $q \in \{1, \dots, Q\}$, and $\beta_1 = 0$. The first of these takes alternative 1 as the reference alternative, as in [Magnac \(2000\)](#), and together with $A_{1i} = 0$ handles the non-identification of type (i). The second takes lagged outcome 1 as the reference lagged state and handles the non-identification of type (ii). The third normalization handles type (iii).

Of course, the baseline linear specification in (3) is only one possible parametrization of $z_t(\cdot, \cdot, \cdot, \theta)$. The framework allows for more flexible choices. For example, the parameters γ_{qr} and β_q in (3) could depend on t , or $Y_{i,t-1}$ and X_{it} could be interacted. Conversely, in some applications a more parsimonious specification may be desirable. In our empirical application, we consider

$$z_t(q, Y_{i,t-1}, X_i, \theta) = \gamma_q \mathbb{1}\{Y_{i,t-1} = q\} + X_{it,q} \beta \quad (4)$$

where the Q coefficients γ_q reward habit formation, while the effect β of the alternative-specific covariate $X_{it,q}$ (i.e price of good q) is restricted to be constant across alternatives.² This specification does not require the normalizations imposed above, since the parameter vector is $\theta = (\gamma_1, \dots, \gamma_Q, \beta)$. We discuss this model further in Sections 5 and 6.

Let $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})$. We assume cross-sectional sampling, that is, (Y_i, Y_{i0}, X_i, A_i) are i.i.d. distributed across i . For convenience, we drop the subscript i from the notation when no confusion can arise. Then, the model assumptions imply that

$$P(Y_t = q | Y^{t-1}, X, A) = \frac{\exp(z_t(q, Y_{t-1}, X, \theta) + A_q)}{\sum_{r=1}^Q \exp(z_t(r, Y_{t-1}, X, \theta) + A_r)}, \quad (5)$$

where $Y^{t-1} = (Y_{t-1}, Y_{t-2}, \dots, Y_0)$ denotes past outcomes. It is convenient to introduce the

²Specification (4) can be viewed as a special case of (3) where the cross-state dependence parameters satisfy $\gamma_{qr} = 0$ for $q \neq r$, and the alternative-specific slopes are restricted to $\beta_q = \beta \iota_q$, where ι_q is the q th canonical basis vector. Hence $X'_{it}\beta_q = X_{it,q}\beta$, explaining the notational change between (3) and (4).

shorthand notation³

$$p_t(q \mid y_{t-1}, x, \theta, \alpha) := \frac{\exp(z_t(q, y_{t-1}, x, \theta) + \alpha_q)}{\sum_{r=1}^Q \exp(z_t(r, y_{t-1}, x, \theta) + \alpha_r)} \quad (6)$$

for the conditional probability of choosing alternative q in period t , given that the previous choice was y_{t-1} . With this notation, the joint probability of observing a sequence of outcomes $y = (y_1, \dots, y_T)$ given the initial condition and fixed effects is

$$p(y \mid y_0, x, \theta, \alpha) := P(Y = y \mid Y_0 = y_0, X = x, A = \alpha) = \prod_{t=1}^T p_t(y_t \mid y_{t-1}, x, \theta, \alpha). \quad (7)$$

The true model parameters are denoted by θ_0 . Throughout, we assume that the model in (7) is correctly specified and all probabilistic statements below are for the model distribution generated under θ_0 .

Our goal is to estimate the common parameters θ_0 from the observed data as $n \rightarrow \infty$ with T fixed. The individual-specific fixed effects A_i are treated as nuisance parameters. Their distribution is left unspecified, and they are allowed to be arbitrarily correlated with the regressors X_i and the initial conditions Y_{i0} . In this setting, treating the individual-specific fixed effects, A_i , as parameters to be estimated will lead to the incidental parameter problem (Neyman and Scott, 1948).

3 Moment conditions free of fixed effects

3.1 Moment condition approach

A classic approach to the incidental parameter problem in parametric models is to look for sufficient statistics for the fixed effects. Conditional on these sufficient statistics, the likelihood does not depend on the fixed effects and can be used for estimation. This approach

³Regarding notation, uppercase X and A denote random vectors, whereas lowercase x and α denote their possible non-random realizations. The object $p_t(q \mid y_{t-1}, x, \theta, \alpha)$ is defined as an ordinary function, so its arguments are non-random. By contrast, stochastic operators such as P and $\mathbb{E}$ act on random quantities and therefore involve X and A .

was pioneered by [Rasch \(1960\)](#) and [Andersen \(1970\)](#). For the static multinomial logit model without regressors, [Chamberlain \(1980\)](#) showed that the vector of choice frequencies serves as a sufficient statistic. For dynamic models, [Magnac \(2000\)](#) and [Aguirregabiria, Gu and Luo \(2021\)](#) have developed sufficient statistic approaches. However, these methods do not generally extend to dynamic multinomial logit models with strictly exogenous regressors.

An alternative approach is to construct moment conditions that depend on the common parameters but not on the individual-specific fixed effects. [Bonhomme \(2012\)](#) developed a general framework for obtaining such moment conditions via “functional differencing.” This approach has been applied to dynamic binary logit models by [Kitazawa \(2022\)](#) and [Honoré and Weidner \(2025\)](#), and to dynamic ordered logit models by [Honoré, Muris and Weidner \(2025\)](#). [Dano \(2023\)](#) provides a general treatment using transition probabilities. The present paper extends this approach to the dynamic multinomial logit model.

Formally, we seek moment functions $m_{y_0}(y, x, \theta) \in \mathbb{R}$ such that

$$\mathbb{E} \left[m_{Y_0}(Y, X, \theta_0) \mid Y_0, X, A \right] = 0 \quad (8)$$

for all possible values of the fixed effect A . Conditioning on the entire covariate path X here uses the strict exogeneity assumption from [Section 2](#), which rules out feedback from past outcomes to future covariates. For given values of $Y_0 = y_0$, $X = x$, θ , and $A = \alpha$, the distribution of $Y = (Y_1, \dots, Y_T) \in \{1, \dots, Q\}^T$ is given by [\(7\)](#). The conditional moment condition [\(8\)](#) can therefore be written as

$$\sum_{y \in \{1, \dots, Q\}^T} m_{y_0}(y, x, \theta) p(y \mid y_0, x, \theta, \alpha) = 0, \quad \text{for all } \alpha \in \mathbb{R}^Q. \quad (9)$$

This is a system of linear equations in the Q^T unknowns $m_{y_0}(y, x, \theta)$, where y ranges over $\{1, \dots, Q\}^T$. The system must hold for all values of α , which imposes infinitely many constraints. The existence of solutions to this infinite-dimensional system is not guaranteed for arbitrary models, and the distributional assumption on the errors is important for finding valid moment conditions. This linear-algebra view of the problem is the starting point of the functional differencing approach of [Bonhomme \(2012\)](#): valid moment functions are those

that lie in the joint null space of the constraints generated by all $\alpha \in \mathbb{R}^Q$.

If we can establish the conditional moment condition (8), then by the law of iterated expectations, we also have the unconditional moment conditions

$$\mathbb{E} [h(Y_0, X, \theta_0) m_{Y_0}(Y, X, \theta_0)] = 0 \quad (10)$$

for any function h such that the expectation is well-defined. Subject to regularity conditions, these unconditional moment conditions can then be used to estimate the model parameters θ_0 by the generalized method of moments (GMM). This approach solves the incidental parameter problem (Neyman and Scott, 1948), because the moment condition (10) does not feature the individual-specific effects A at all. No assumptions are imposed on the distribution of those nuisance parameters, and they need not be estimated.

Finding a non-trivial solution $m_{y_0}(y, x, \theta)$ to the system (9) is generally hard. The system must hold for all $\alpha \in \mathbb{R}^Q$, which imposes infinitely many restrictions on the Q^T unknowns. For arbitrary parametric models, only the trivial solution $m_{y_0} \equiv 0$ may exist. Even when non-trivial solutions do exist, they are typically not available in closed form. Bonhomme (2012) provides an abstract characterization of the space of valid moment conditions through functional differencing, but this characterization does not by itself deliver explicit moment functions that can be used for estimation. As a result, explicit constructions of moment functions free of the fixed effects have so far been carried out on a model-by-model basis, relying on features specific to each setting, such as the logit structure in the binary and ordered choice models. The main contribution of this paper is to provide a similar explicit construction for the dynamic multinomial logit model, using the transformed model approach developed in the next subsection.

3.2 Transformed model

Our construction of valid moment functions is based on the concept of a transformed model. The key observation is that the fixed effects A enter the model probabilities in a specific way, and by comparing the true model with a suitably constructed transformed model, one can

eliminate the dependence on A .

The intuition behind this construction is as follows. The fixed effects A enter the joint probability $p(y \mid y_0, x, \theta, \alpha)$ in two ways: (i) through the product $\prod_{t=1}^T \exp(\alpha_{y_t})$ in the numerators of the period-by-period choice probabilities, and (ii) through their denominators $\prod_{t=1}^T \sum_{r=1}^Q \exp(z_t(r, y_{t-1}, x, \theta) + \alpha_r)$. Crucially, the numerator product depends on the outcome sequence only through the unordered multiset $\{y_1, \dots, y_T\}$, which renders it invariant under any permutation σ of the time indices. This implies that comparing the true model with a transformed one that reshuffles outcomes while preserving the denominators will eliminate the fixed effect contribution. We will show below that the resulting likelihood ratio $p^\sigma(y \mid y_0, x, \theta, \alpha)/p(y \mid y_0, x, \theta, \alpha)$ is indeed free of α (Lemma 2), and that when σ is a cyclic permutation this ratio can be turned into valid moment conditions for estimation.

Let $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ be a permutation of the time indices. We define the transformed model probability as

$$p^\sigma(y \mid y_0, x, \theta, \alpha) := p_1(y_{\sigma(T)} \mid y_0, x, \theta, \alpha) \prod_{t=2}^T p_t(y_{\sigma(t-1)} \mid y_{t-1}, x, \theta, \alpha). \quad (11)$$

Compared to the true model, where period t generates outcome y_t , the transformed model reassigns which outcome is generated in each period: period 1 generates $y_{\sigma(T)}$, and period $t \geq 2$ generates $y_{\sigma(t-1)}$. The true model is the special case $\sigma = \sigma_+$, where σ_+ is the cyclic increment $\sigma_+(t) = t + 1$ for $t < T$ and $\sigma_+(T) = 1$.

The expression p^σ in (11) is nonnegative by construction, but it does not sum to one over all $y \in \{1, \dots, Q\}^T$ for arbitrary permutations σ . For a general permutation σ , the factors in (11) do not necessarily generate a single cyclic T -step Markov chain through all T outcomes. To see this formally, define

$$\tau(k) := \sigma^k(T), \quad k = 1, \dots, T, \quad (12)$$

where σ^k is the k -fold composition of σ , e.g. $\sigma^2(T) = \sigma(\sigma(T))$. The transformed model has a proper Markov structure only if τ visits each index in $\{1, \dots, T\}$ exactly once. In this case,

τ is itself a permutation of $\{1, \dots, T\}$, and p^σ can be rewritten as

$$p^\sigma(y \mid y_0, x, \theta, \alpha) = p_1(y_{\tau(1)} \mid y_0, x, \theta, \alpha) \prod_{k=2}^T p_{\tau(k-1)+1}(y_{\tau(k)} \mid y_{\tau(k-1)}, x, \theta, \alpha). \quad (13)$$

We call the sequence $\tau = (\tau(1), \tau(2), \dots, \tau(T))$ the generation order of σ . The name reflects the role that τ plays in the representation (13): starting from y_0 , the transformed model generates $y_{\tau(1)}$ first, then $y_{\tau(2)}$, and so on, with $y_{\tau(T)}$ generated last. In other words, the generation order of σ is the order in which the outcomes $y_1, \dots, y_T$ are produced by the transformed model p^σ as a Markov chain. It turns out that (11) can be rewritten as (13) if and only if σ is a *cyclic* permutation of $\{1, \dots, T\}$. This is a well-known concept in combinatorics, which simply means that $T, \sigma(T), \sigma^2(T), \sigma^3(T), \dots, \sigma^{T-1}(T)$ are all distinct.⁴

To see why cyclic permutations are the relevant class, recall that every permutation σ of $\{1, \dots, T\}$ decomposes into disjoint cycles, and that the sequence $\tau(k) = \sigma^k(T)$ traces only the cycle of σ containing T . The reformulation (13) represents the transformed model as a single Markov chain through all T time periods, which is possible only when the cycle of T has length T , that is, when σ is cyclic. For a non-cyclic σ , the cycle of T has length strictly less than T , and $\tau(k)$ returns to T before visiting the time indices in the other cycles of σ . In this case p^σ factors as a product of independent Markov chains, one per cycle of σ . The chains associated with cycles not containing T are “closed loops” that start and end at the same outcome, and for generic (x, θ, α) they do not sum to one over the corresponding outcomes. The cyclic case is therefore the only one for which $\sum_y p^\sigma(y \mid y_0, x, \theta, \alpha) = 1$ holds for all (x, θ, α) ; see Appendix B for the factorization in the non-cyclic case.

Once we have the reformulation (13), then it is straightforward to show that $p^\sigma(y \mid y_0, x, \theta, \alpha)$ sums to one, as summarized by the following lemma.

Lemma 1 *If $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ is a cyclic permutation, then*

$$\sum_{y \in \{1, \dots, T\}^T} p^\sigma(y \mid y_0, x, \theta, \alpha) = 1. \quad (14)$$

⁴Formally, a permutation $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ is cyclic if and only if the orbit of T under σ has length T . Equivalently, $\{\sigma^k(T) : k = 0, 1, \dots, T-1\} = \{1, \dots, T\}$.

Proof of Lemma 1. Since σ is cyclic, the sequence $\tau(k) = \sigma^k(T)$ visits each index in $\{1, \dots, T\}$ exactly once. Hence (11) is a T -step Markov chain written in a different time order, see (13). Therefore iterated summation gives (14) since at each step $\sum_{j=1}^Q p_t(j \mid \cdot) = 1$.

■

3.3 Moment construction

Let $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ be a permutation. Define $\sigma(0) := \sigma(T)$. Then (11) can be written more concisely as

$$p^\sigma(y \mid y_0, x, \theta, \alpha) = \prod_{t=1}^T p_t(y_{\sigma(t-1)} \mid y_{t-1}, x, \theta, \alpha), \quad (15)$$

and this definition of $\sigma(0)$ will simplify many expressions below.

The ratio of the transformed model probability to the true model probability plays a central role in our construction. The following lemma shows that the ratio $p^\sigma(y \mid y_0, x, \theta, \alpha)/p(y \mid y_0, x, \theta, \alpha)$ is free of α and equal to

$$S_{y_0}^\sigma(y, x, \theta) := \exp\left(\sum_{t=1}^T \left[z_t(y_{\sigma(t-1)}, y_{t-1}, x, \theta) - z_t(y_t, y_{t-1}, x, \theta) \right]\right). \quad (16)$$

Lemma 2 *For any permutation $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ we have*

$$\frac{p^\sigma(y \mid y_0, x, \theta, \alpha)}{p(y \mid y_0, x, \theta, \alpha)} = S_{y_0}^\sigma(y, x, \theta), \quad (17)$$

implying that this ratio is free of α .

Proof of Lemma 2. Using (15) and the definition of the true model,

$$\frac{p^\sigma(y \mid y_0, x, \theta, \alpha)}{p(y \mid y_0, x, \theta, \alpha)} = \prod_{t=1}^T \frac{p_t(y_{\sigma(t-1)} \mid y_{t-1}, x, \theta, \alpha)}{p_t(y_t \mid y_{t-1}, x, \theta, \alpha)}.$$

For each t , the normalizing denominators cancel, so

$$\frac{p_t(y_{\sigma(t-1)} \mid y_{t-1}, x, \theta, \alpha)}{p_t(y_t \mid y_{t-1}, x, \theta, \alpha)} = \exp\left(z_t(y_{\sigma(t-1)}, y_{t-1}, x, \theta) - z_t(y_t, y_{t-1}, x, \theta)\right) \exp\left(\alpha_{y_{\sigma(t-1)}} - \alpha_{y_t}\right).$$

Multiplying over t gives $S_{y_0}^\sigma(y, x, \theta)$ times

$$\exp\left(\sum_{t=1}^T \alpha_{y_{\sigma(t-1)}} - \sum_{t=1}^T \alpha_{y_t}\right).$$

Since σ is a permutation and $\sigma(0) = \sigma(T)$, the indices $\sigma(t-1)$, $t = 1, \dots, T$, are a reordering of $1, \dots, T$. Hence $\sum_{t=1}^T \alpha_{y_{\sigma(t-1)}} = \sum_{t=1}^T \alpha_{y_t}$, so the α term cancels. This yields (17). ■

Next, let σ be a cyclic permutation. Then the simplest valid moment condition that follows directly from Lemma 1 and Lemma 2 is

$$\mathbb{E}\left[1 - S_{Y_0}^\sigma(Y, X, \theta_0) \mid Y_0, X, A\right] = 0. \quad (18)$$

To see this, note that

$$\mathbb{E}\left[S_{Y_0}^\sigma(Y, X, \theta_0) \mid Y_0, X, A\right] = \sum_y S_{Y_0}^\sigma(y, X, \theta_0) p(y \mid Y_0, X, \theta_0, A) = \sum_y p^\sigma(y \mid Y_0, X, \theta_0, A) = 1,$$

where the second equality uses (17) and the last equality uses (14).

While valid, the moment function $1 - S_{y_0}^\sigma(y, x, \theta)$ may not be the most useful in practice, because it typically has $\theta = 0$ as a root regardless of the true parameter value. Indeed, for the linear specification (3) we have $z_t(\cdot, \cdot, x, 0) = 0$, so $S_{y_0}^\sigma(y, x, 0) = 1$ for every permutation and every outcome sequence, and the moment function $1 - S^\sigma$ vanishes identically at $\theta = 0$. Moment conditions based on (18) alone therefore cannot separate the true parameter value from zero. The following theorem therefore presents more general valid moment functions.

The construction works by pairing two cyclic permutations σ_1 and σ_2 whose generation orders τ_1 and τ_2 agree on their first $k-1$ entries, but differ at position k . For both of the transformed models, the first $k-1$ outcomes $Y_{\tau_1(1)}, \dots, Y_{\tau_1(k-1)}$ are produced by identical Markov transitions, so conditioning on these outcomes removes the shared history. The

divergence in position k is what makes the moment function nontrivial. The outcomes in the remaining positions $k+1, \dots, T$ telescope out of the expectation through the Markov factorization (13), leaving an exact cancellation between the two S^σ terms.

Theorem 1 *Let σ_1 and σ_2 be two cyclic permutations of $\{1, \dots, T\}$, and define $\tau_1(k) := \sigma_1^k(T)$ and $\tau_2(k) := \sigma_2^k(T)$ for $k \in \{1, \dots, T\}$, analogous to (12). Assume $\sigma_1 \neq \sigma_2$ and let $k \in \{1, \dots, T\}$ be the smallest integer such that $\tau_1(k) \neq \tau_2(k)$. Then for any $q \in \{1, \dots, Q\}$, the moment function*

$$m_{y_0}^{\sigma_1, \sigma_2, q}(y, x, \theta) := \mathbb{1}\{y_{\tau_1(k)} = q\} S_{y_0}^{\sigma_1}(y, x, \theta) - \mathbb{1}\{y_{\tau_2(k)} = q\} S_{y_0}^{\sigma_2}(y, x, \theta) \quad (19)$$

satisfies

$$\mathbb{E} \left[m_{Y_0}^{\sigma_1, \sigma_2, q}(Y, X, \theta_0) \mid Y_{\tau_1(1)}, Y_{\tau_1(2)}, \dots, Y_{\tau_1(k-1)}, Y_0, X, A \right] = 0. \quad (20)$$

This implies that for any function $g : \{1, \dots, Q\}^{k-1} \rightarrow \mathbb{R}$ we have

$$\mathbb{E} \left[g(Y_{\tau_1(1)}, Y_{\tau_1(2)}, \dots, Y_{\tau_1(k-1)}) m_{Y_0}^{\sigma_1, \sigma_2, q}(Y, X, \theta_0) \mid Y_0, X, A \right] = 0. \quad (21)$$

Proof of Theorem 1. It suffices to prove (20), since (21) follows by multiplying by $g(Y_{\tau_1(1)}, \dots, Y_{\tau_1(k-1)})$ and taking iterated expectations.

Fix arbitrary $q_1, \dots, q_{k-1} \in \{1, \dots, Q\}$. For $i \in \{1, 2\}$, Lemma 2 gives $S^{\sigma_i} = p^{\sigma_i}/p$, so

$$\begin{aligned} & \mathbb{E} \left[\mathbb{1}\{Y_{\tau_i(k)} = q\} S_{Y_0}^{\sigma_i}(Y, X, \theta_0) \cdot \prod_{j=1}^{k-1} \mathbb{1}\{Y_{\tau_1(j)} = q_j\} \mid Y_0, X, A \right] \\ &= \sum_{\substack{y \in \{1, \dots, Q\}^T: \\ y_{\tau_1(j)} = q_j \ \forall j=1, \dots, k-1}} \mathbb{1}\{y_{\tau_i(k)} = q\} p^{\sigma_i}(y \mid Y_0, X, \theta_0, A). \end{aligned}$$

By the Markov factorization (13), summing over the remaining components $y_{\tau_i(k+1)}, \dots, y_{\tau_i(T)}$

telescopes to one, and the right-hand side equals

$$p_1(q_1 \mid Y_0, X, \theta_0, A) \prod_{j=2}^{k-1} p_{\tau_i(j-1)+1}(q_j \mid q_{j-1}, X, \theta_0, A) \cdot p_{\tau_i(k-1)+1}(q \mid q_{k-1}, X, \theta_0, A).$$

For $k = 1$ no outcomes are held fixed, and the right-hand side is simply $p_1(q \mid Y_0, X, \theta_0, A)$. Since $\tau_1(j) = \tau_2(j)$ for all $j \leq k - 1$, this expression is the same for $i = 1$ and $i = 2$. Hence the two terms in (19) cancel. Since this holds for all $(q_1, \dots, q_{k-1})$, equation (20) follows. ■

3.3.1 Example: moment conditions for $T = 3$

To illustrate, consider $T = 3$. There are $(T - 1)! = 2$ cyclic permutations of $\{1, 2, 3\}$: the cyclic increment σ_+ , which gives the true model, and the reverse cycle σ_1 defined by $\sigma_1(1) = 3, \sigma_1(2) = 1, \sigma_1(3) = 2$. Taking $\sigma_2 = \sigma_+$, the generation orders are $\tau_1 = (2, 1, 3)$ and $\tau_2 = (1, 2, 3)$, which first differ at $k = 1$. Since $\sigma_2 = \sigma_+$ is the true model, $S_{y_0}^{\sigma_2}(y, x, \theta) = 1$, and Theorem 1 yields Q conditional moment conditions:

$$\mathbb{E} \left[\mathbb{1}\{Y_2 = q\} S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - \mathbb{1}\{Y_1 = q\} \mid Y_0, X, A \right] = 0, \quad q = 1, \dots, Q, \quad (22)$$

where, by (16),

$$\begin{aligned} S_{y_0}^{\sigma_1}(y, x, \theta) = & \exp \left(z_1(y_2, y_0, x, \theta) - z_1(y_1, y_0, x, \theta) \right. \\ & + z_2(y_3, y_1, x, \theta) - z_2(y_2, y_1, x, \theta) \\ & \left. + z_3(y_1, y_2, x, \theta) - z_3(y_3, y_2, x, \theta) \right). \end{aligned} \quad (23)$$

For $Q = 2$, the moment conditions (22) recover the leading result for the dynamic binary logit model with $T = 3$ in Honoré and Weidner (2025). The transformed model approach provides a particularly transparent derivation of these known moment conditions for $Q = 2$.

More generally, varying $(\sigma_1, \sigma_2, q, g)$ over all admissible choices in Theorem 1 generates a large family of valid moment functions, many of which are linearly dependent. Determining the number of linearly independent moment conditions that this construction yields, as a

function of Q and T , is the subject of the next subsection.

3.4 Counting of linearly independent moment conditions

Theorem 1 produces a large family of moment functions as we vary the cyclic permutations σ_1, σ_2 and the weighting function g . Many choices will generate moment conditions that are redundant, in the sense that they can be written as linear combinations of others in the family. It is therefore useful to know how many linearly independent moment conditions can be obtained from the theorem.

We now describe an explicit parameterization that extracts $\ell(Q, T)$ moment functions from Theorem 1, where $\ell(Q, T)$ is the count given in (25) below. The moment functions will be indexed by tuples $(j, \mathbf{t}, s_2, \mathbf{q})$ where $j \in \{0, 1, \dots, T-3\}$ is an integer that records the number of shared time indices between the two generation orders of Theorem 1, $\mathbf{t} = (t_1 < \dots < t_j)$ is a collection of j time periods in $\{1, \dots, T-1\}$, $s_1 := \min(\{1, \dots, T-1\} \setminus \mathbf{t})$ is therefore determined by $\mathbf{t}$, and s_2 is any element of $\{1, \dots, T-1\} \setminus (\mathbf{t} \cup \{s_1\})$, and $\mathbf{q} = (q_0, q_1, \dots, q_j) \in \{1, \dots, Q\}^{j+1}$ is a vector of outcome values. For each such tuple the corresponding moment function takes the form

$$m_{y_0}^{j, \mathbf{t}, s_1, s_2, \mathbf{q}}(y, x, \theta) := \left[\prod_{l=1}^j \mathbb{1}\{y_{t_l} = q_l\} \right] \left[\mathbb{1}\{y_{s_1} = q_0\} S_{y_0}^{\sigma_1}(y, x, \theta) - \mathbb{1}\{y_{s_2} = q_0\} S_{y_0}^{\sigma_2}(y, x, \theta) \right], \quad (24)$$

where σ_1 and σ_2 are the two cyclic permutations of $\{1, \dots, T\}$ whose generation orders agree on the first j entries $(t_1, \dots, t_j)$ but differ at position $j+1$, with σ_1 placing s_1 at that position and σ_2 placing s_2 .⁵ Since the generation orders first differ at position $k = j+1$, Theorem 1 applies with this k , $q = q_0$, and weighting function $g(y_{t_1}, \dots, y_{t_j}) = \prod_{l=1}^j \mathbb{1}\{y_{t_l} = q_l\}$, and guarantees that $m_{y_0}^{j, \mathbf{t}, s_1, s_2, \mathbf{q}}$ satisfies (8) at $\theta = \theta_0$.

To avoid counting redundant moment functions, we further impose the restriction $t_r = r$

⁵Explicitly, σ_1 has generation order $\tau_1 = (t_1, \dots, t_j, s_1, r_1, \dots, r_{T-j-3}, s_2, T)$ and σ_2 has generation order $\tau_2 = (t_1, \dots, t_j, s_2, r_1, \dots, r_{T-j-3}, s_1, T)$, where $r_1 < \dots < r_{T-j-3}$ are the remaining elements of $\{1, \dots, T-1\} \setminus (\mathbf{t} \cup \{s_1, s_2\})$ in increasing order.

for $r = 1, \dots, \max(0, j - (Q - 2))$ on $\mathbf{t}$.⁶

Counting the distinct tuples $(j, \mathbf{t}, s_2, \mathbf{q})$ that satisfy all the restrictions above then gives⁷

$$\ell(Q, T) = \sum_{j=0}^{T-3} \binom{T-1-\max(j-(Q-2), 0)}{\min(Q-2, j)} \cdot (T-2-j) \cdot Q^{j+1}. \quad (25)$$

For small values of Q this simplifies to $\ell(2, T) = 2^T - 2T$, $\ell(3, T) = 3^T - \binom{3T}{2} + 6(T-1)$, and $\ell(4, T) = 4^T - \binom{4T}{3} + 40(T-1)^2$. More generally, for $Q \geq T-1$, it becomes $\ell(Q, T) = Q^T - Q(2+Q-T)(1+Q)^{T-2}$. In each case the formula has the structure $\ell(Q, T) = Q^T - r(Q, T)$, where the ambient dimension Q^T counts all real-valued functions of the outcome sequence y , and the subtracted term $r(Q, T)$ counts the linear restrictions on m imposed by requiring the moment condition (9) to hold for all values of the fixed effect α .

Numerical investigations and the partial analytical results discussed in Appendix A lead us to the following conjecture.

Conjecture 1 *For $T \geq 3$ and $Q \geq 2$ and for generic (y_0, x, θ_0) , the $\ell(Q, T)$ moment functions in (24) are linearly independent, and every valid moment function of the form (8) at $\theta = \theta_0$ is a linear combination of them.*

The status of the conjecture is as follows. For $Q = 2$ the lower bound follows from the explicit moment functions of Honoré and Weidner (2025), and the matching upper bound is established in subsequent work; see Dobronyi, Gu, Kim and Russell (2021) and Dano (2023). For $T = 3$ and any $Q \geq 2$, we prove in Appendix A that the $\ell(Q, 3) = Q$ moment functions in (24) are linearly independent, so the dimension of the space of valid moment conditions is at least $\ell(Q, 3)$. For all combinations of (Q, T) reported in Table 1, the conjecture has been verified by direct numerical computation. A general analytical proof for $Q \geq 3$ and $T \geq 4$ remains an open problem. Appendix A discusses two natural proof strategies (a polynomial

⁶Without this restriction the construction would yield more tuples than $\ell(Q, T)$, but those additional moment functions would be linearly dependent on those in the constrained set. For $j \leq Q-2$ the restriction is vacuous; for $j > Q-2$ it fixes the first $j-(Q-2)$ entries of $\mathbf{t}$ to the smallest possible values $1, 2, \dots, j-(Q-2)$.

⁷For each j , the restriction on $\mathbf{t}$ fixes $\max(0, j-(Q-2))$ of its entries and leaves $\min(j, Q-2)$ entries to be chosen freely from the $T-1-\max(0, j-(Q-2))$ remaining positions in $\{1, \dots, T-1\}$, giving $\binom{T-1-\max(j-(Q-2), 0)}{\min(Q-2, j)}$ choices for $\mathbf{t}$. Given $\mathbf{t}$, there are $T-2-j$ choices for $s_2 \in \{1, \dots, T-1\} \setminus (\mathbf{t} \cup \{s_1\})$, and Q^{j+1} choices for $\mathbf{q}$. Summing the product of these three counts over $j \in \{0, 1, \dots, T-3\}$ yields (25).

	$T = 1$	$T = 2$	$T = 3$	$T = 4$	$T = 5$	$T = 6$	$T = 7$	$T = 8$
$Q = 2$	0	0	2	8	22	52	114	240
$Q = 3$	0	0	3	33	162	606	2013	6327
$Q = 4$	0	0	4	56	524	3072	14548	62536
$Q = 5$	0	0	5	85	965	9145	62500	358280
$Q = 6$	0	0	6	120	1602	17844	179094	1399680
$Q = 7$	0	0	7	161	2471	31633	364791	3929793
$Q = 8$	0	0	8	208	3608	52192	679976	8274160

Table 1: Conjectured number of linearly independent moment conditions $\ell(Q, T)$ for various values of Q (the number of alternatives) and T (the number of time periods after the initial condition). Verified by numerical computation for all entries in the table.

expansion of the likelihood and a direct linear-independence argument on the moments in (24)) and identifies the technical obstacle.

The qualifier “generic” in the conjecture means that it is expected to hold outside a measure-zero subset of (Y_0, X, θ_0) . For special values of X and θ_0 , additional moment conditions may become available, so the total number of linearly independent moment conditions can exceed $\ell(Q, T)$.

Table 1 reports $\ell(Q, T)$ for combinations relevant for empirical applications. Two features are worth noting. First, $\ell(Q, T) = 0$ for $T \leq 2$ for every Q , which confirms that at least three time periods after the initial condition are required for moment conditions free of the fixed effects to exist for generic parameters. Second, the count grows quickly in both Q and T : with $Q = 4$ alternatives and $T = 5$ time periods, the parameterization yields 524 linearly independent moment conditions for each value of the initial condition. Only a small subset of these moment conditions is typically used for estimation in practice. We return to this issue in Section 6.

4 Moment conditions for more general models

Most of the derivations in Sections 3.2 and 3.3 are generic and do not use the specific structure of the multinomial logit model. The only exception is Lemma 2 which establishes that for

any permutation $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$,

$$S_{y_0}^\sigma(y, x, \theta) := \frac{p^\sigma(y \mid y_0, x, \theta, \alpha)}{p(y \mid y_0, x, \theta, \alpha)} \quad (26)$$

is free of α . In this section, we characterize a larger class of models for which this is true, and it follows that moment conditions can be constructed using the arguments of Sections 3.2 and 3.3 for any model in this larger class.

Consider the model for the joint distribution of $(Y_1, Y_2, \dots, Y_T)$:

$$p(y \mid y_0, x, \theta, \alpha) = P\left(Y = y \mid Y_0 = y_0, x, A = \alpha\right) = \prod_{t=1}^T f(y_t, z_t(y_{t-1}, x, \theta), \alpha) \quad (27)$$

for functions f and $z_1, \dots, z_T$.

Lemma 3 *Let p be defined as in (27). Assume that for every $(\eta_1, \dots, \eta_T)$ in the range of the functions z_t , the joint distribution*

$$\prod_{t=1}^T f(y_t, \eta_t, \alpha)$$

on the outcome space of $(Y_1, \dots, Y_T)$ admits a statistic $S = s(Y_1, \dots, Y_T)$ such that the conditional distribution of $(Y_1, \dots, Y_T)$ given S does not depend on α , and that s is symmetric in its arguments (so that $s(y_{\pi(1)}, \dots, y_{\pi(T)}) = s(y_1, \dots, y_T)$ for every permutation π of $\{1, \dots, T\}$). Then for any permutation $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$, the ratio

$$\frac{p^\sigma(y \mid y_0, x, \theta, \alpha)}{p(y \mid y_0, x, \theta, \alpha)} = S_{y_0}^\sigma(y, x, \theta)$$

is free of α .

Proof of Lemma 3. The assumptions imply that for any values of $\eta_1, \dots, \eta_T$, the ratio $\frac{\prod_{t=1}^T f(s_t, \eta_t, \alpha)}{\prod_{t=1}^T f(y_t, \eta_t, \alpha)}$ does not depend on α provided that $(s_1, \dots, s_T)$ is a permutation of $(y_1, \dots, y_T)$. With $\eta_t = z_t(y_{t-1}, x, \theta)$, we then have

$$\frac{p^\sigma(y \mid y_0, x, \theta, \alpha)}{p(y \mid y_0, x, \theta, \alpha)} = \frac{f(y_{\sigma(T)}, \eta_1, \alpha) \prod_{t=2}^T f(y_{\sigma(t-1)}, \eta_t, \alpha)}{\prod_{t=1}^T f(y_t, \eta_t, \alpha)}.$$

Since $(y_{\sigma(T)}, y_{\sigma(1)}, \dots, y_{\sigma(T-1)})$ is a permutation of $(y_1, \dots, y_T)$, this does not depend on α .

■

Combining Lemma 3 with the results in Sections 3.2 and 3.3 implies that for any static model for which there exists a symmetric sufficient statistic for A , it is possible to construct moment conditions for the corresponding dynamic model in which Y_{t-1} is an explanatory variable in period t . This unifies some known results and makes it clear how to construct moment conditions in models that have not previously been considered in the literature. Moreover, the logic behind the construction also applies to continuous distributions. We provide some examples below.

4.1 Example: exponential–family panel models (Hahn 1997)

A rich class of models satisfying the hypotheses of Lemma 3 is provided by the exponential–family panel models studied by Hahn (1997). In the notation of (27), the period– t density takes the form

$$f(y_t, \eta, \alpha) = \exp\left(s(y_t)' \alpha e(\theta) + d(\eta, \alpha, \theta) + b(y_t, \eta, \theta)\right), \quad (28)$$

where α is the fixed effect, $\eta = z_t(y_{t-1}, x, \theta)$ is the index, $s(\cdot)$ is a vector of canonical statistics, $e(\theta)$ is a known function of the common parameter, $d(\cdot)$ is the log–normalizing constant (which does not involve y_t), and $b(\cdot)$ collects the remaining terms, which involve y_t and the index but not the fixed effect. This is the model in equation (1) of Hahn (1997), with the role of the regressors played here by the index η , and as Hahn (1997) notes, this class contains the panel linear regression model with normal errors, the panel logit model, the panel Poisson regression model, and the panel negative binomial model.

Consider the static case where $\eta = z_t(y_{t-1}, x, \theta)$ does not depend on y_{t-1} . In this case,

$$S(Y) = \sum_{t=1}^T s(Y_t)$$

is a sufficient statistic for α , and it is symmetric in $(Y_1, \dots, Y_T)$ (Lemma 4 of Appendix C).

The assumptions of Lemma 3 are therefore satisfied, and the moment–condition construction of Sections 3.2–3.3 applies to a model that includes lagged dependent variables as explanatory variables.

To obtain the moment functions explicitly, form the ratio p^σ/p using (28). The normalizing constants $d(\eta_t, \alpha, \theta)$ cancel period by period; and because $\sum_t s(y_{\sigma(t-1)}) = \sum_t s(y_t)$, the fixed effect term $s(y_t)' \alpha e(\theta)$ cancels as well. What remains is

$$S_{y_0}^\sigma(y, x, \theta) = \exp\left(\sum_{t=1}^T [b(y_{\sigma(t-1)}, \eta_t, \theta) - b(y_t, \eta_t, \theta)]\right), \quad \eta_t = z_t(y_{t-1}, x, \theta), \quad (29)$$

which is free of α , in agreement with Lemma 3. Exactly as in Theorem 1, take $T = 3$, $\sigma_2 = \sigma_+$ (so that $S^{\sigma_2} \equiv 1$), and σ_1 the reverse cycle; when Y is continuous the indicator $\mathbb{1}\{Y_{\tau(k)} = q\}$ is replaced by an arbitrary bounded function g , and the moment condition is that, for every such g ,

$$\mathbb{E}\left[g(Y_2) S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - g(Y_1) \mid Y_0, X, A\right] = 0. \quad (30)$$

The validity of (30) is established more formally in Proposition 3 of Appendix C.

4.2 Example: sufficient statistics that depend on θ

The model in (28) implies a sufficient statistic that is free of the common parameter and the regressors. However, the construction of moment conditions suggested above requires only that the static model admits a sufficient statistic for A that is symmetric in the time indices. It therefore applies more broadly, including cases in which the relevant statistic depends on θ , so long as symmetry across t is preserved. A bivariate gamma panel model in the spirit of Lindsay (1982) illustrates this. For each t the outcome is the pair $Y_t = (Y_{1t}, Y_{2t})$ where, conditional on the past, the regressors, and the fixed effect A , the two components

are independent with⁸

$$Y_{1t} \sim \text{Gamma}(z_{1t}, A), \quad Y_{2t} \sim \text{Gamma}(z_{2t}, \phi A), \quad (31)$$

where $z_{jt} = z_{jt}(y_{t-1}, X, \theta)$ for $j = 1, 2$ and $\theta = (\beta'_1, \beta'_2, \phi)'$. In the static case where the $z_{jt}(y_{t-1}, X, \theta)$ do not depend on y_{t-1} ,

$$s_\theta(Y) = \sum_{t=1}^T Y_{1t} + \phi \sum_{t=1}^T Y_{2t}$$

is a sufficient statistic for A in the sense that the distribution of $(Y_1, \dots, Y_T)$ conditional on $s_\theta(Y)$ does not depend on A . It is symmetric in the time indices, but in contrast to (28), it depends on the common parameter θ through ϕ , so the model lies outside the class of [Hahn \(1997\)](#). Symmetry is nonetheless all that our moment conditions construction requires, and the same cancellation gives

$$S_{y_0}^\sigma(y, X, \theta) = \exp\left(\sum_{t=1}^T [z_{1t}(\log y_{1,\sigma(t-1)} - \log y_{1t}) + z_{2t}(\log y_{2,\sigma(t-1)} - \log y_{2t})]\right), \quad (32)$$

free of A , so that, for any bounded g on the outcome space,

$$\mathbb{E}\left[g(Y_2) S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - g(Y_1) \mid Y_0, X, A\right] = 0; \quad (33)$$

see Appendix C for details. The cross-equation parameter ϕ enters only through the fixed effect-coupled rate of Y_{2t} and hence cancels from S^σ together with A . The construction thus delivers moment conditions for the index parameters (β_1, β_2) , while identifying ϕ would require other arguments.⁹ Of course in this case, different moment conditions for (β_1, β_2) could have also been constructed using the marginal distributions for Y_{1t} and Y_{2t} and the fact that the fixed effect enters the means of Y_{1t} and Y_{2t} multiplicatively.

⁸These use the common shape-rate parameterization.

⁹For example, ϕ may be recovered as $\phi = \mathbb{E}\left[\frac{Y_{1t}}{z_{1t}}\right] / \mathbb{E}\left[\frac{Y_{2t}}{z_{2t}}\right]$ if β_1, β_2 are identified. This readily follows from the expression of the mean for a Gamma distribution.

4.3 Example: linear panel data models with normal errors

The linear panel data model with normal errors and additive fixed effects,

$$Y_t = z_t(Y_{t-1}, X, \theta) + A + \varepsilon_t, \quad \varepsilon_t \mid (Y^{t-1}, X, A) \sim \mathcal{N}(0, \sigma^2), \quad (34)$$

is also a special case of (28), with the fixed effect entering through the canonical statistic $s(y_t) = y_t$. Hence $S(Y) = \sum_{t=1}^T Y_t$ is a symmetric sufficient statistic for A , and Lemma 3 applies. In forming p^σ/p , the terms in y_t^2 , which do not involve the index, cancel by symmetry along with the fixed effect, leaving

$$S_{y_0}^\sigma(y, x, \theta) = \exp\left(\frac{1}{\sigma^2} \sum_{t=1}^T z_t(y_{t-1}, x, \theta) (y_{\sigma(t-1)} - y_t)\right), \quad (35)$$

and the moment condition $\mathbb{E}[g(Y_2) S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - g(Y_1) \mid Y_0, X, A] = 0$ for any bounded g .

This example is instructive because it exposes the limits of the moment condition construction above. The model (34) already admits an obvious and well-behaved set of fixed effect-free moment conditions: because $Y_t - z_t = A + \varepsilon_t$, any contrast that differences out the fixed effect, such as

$$\mathbb{E}[(Y_t - z_t) - (Y_{t-1} - z_{t-1}) \mid Y_0, X, A] = 0, \quad (36)$$

is linear in the data, possesses moments of all orders, and is the basis for the standard linear panel data estimators (Arellano and Bond, 1991). When z_t is linear in the lagged outcome, the moment function (35) produced by our construction is, by contrast, the exponential of a quadratic¹⁰ form in the Gaussian outcomes — the quadratic arising from the lagged outcome inside z_t — and is consequently heavy-tailed. Its conditional variance need not exist: it is finite only when the state-dependence coefficient lies in a bounded interval, and successively smaller intervals are required for the higher moments,¹¹ and outside these ranges the moment

¹⁰Interestingly, using a different approach Kitazawa (2001) derived moment conditions for a model like (34) that are based on exponential of linear functions of Gaussian outcomes.

¹¹For example, with $z_t(y_{t-1}, x, \theta) = \gamma y_{t-1} + x'_t \beta$ and $T = 3$, the conditional second moment $\mathbb{E}[(S_{Y_0}^{\sigma_1})^2 \mid Y_0, X, A]$ is finite if and only if γ lies between the two real roots of $1 - 4\gamma^2 + 4\gamma^3 - 4\gamma^4$, that is, $\gamma \in (-0.400, 0.575)$; more generally the k -th conditional moment is finite if and only if $2k(k-1)\gamma^2(1-\gamma+\gamma^2) < 1$.

has infinite variance and the associated estimator is no longer $\sqrt{n}$ -asymptotically normal. Even where the variance is finite, the moment is strongly right-skewed, so generalized method of moments estimators based on it can be very imprecise. We conclude from this example that the transformed-model approach can sometimes best be understood as a device for generating valid fixed effect-free moment conditions when no problem-specific reduction is apparent. The moment functions it delivers are always valid, but, as (35) illustrates, they need not be well behaved, and they may be dominated by simple alternatives such as the differencing moments (36) that a model-specific analysis would suggest. The construction therefore complements, but does not substitute for, case-specific reasoning about how best to eliminate the fixed effects.

5 Identification

In earlier work, [Honoré and Kyriazidou \(2000\)](#) studied the dynamic multinomial logit model with the linear specification (3) and showed that θ_0 can be identified with only $T = 3$ periods. The key restriction they require is that the regressors do not vary over the last two periods, $X_2 = X_3$. This makes it possible to construct a conditional likelihood that eliminates the fixed effects and leads to an elegant identification argument.¹²

The drawback is that this restriction is demanding when covariates are continuous, as in our empirical application which features product-specific prices. In that case, the event $X_2 = X_3$ has probability zero, so estimation relies only on the few observations for which X_2 and X_3 are close, implying in turn a slower-than- $\sqrt{n}$ rate of convergence.

This section shows that this shortcoming can be avoided by leveraging the fixed effect free moment condition in Theorem 1. We illustrate this point in the context of our empirical

¹²It can be shown that under the restriction $X_2 = X_3$, model (7) admits additional valid moment functions beyond those presented in Theorem 1. The score of the conditional likelihood in [Honoré and Kyriazidou \(2000\)](#) involves a particular linear combination of these extra moments. Here, our argument relies only on the moment conditions in Theorem 1, which remain valid without imposing $X_2 = X_3$ and can therefore be used for $\sqrt{n}$ -consistent estimation with continuous covariates.

specification (4),

$$z_t(q, Y_{i,t-1}, X_i, \theta) = \gamma_q \mathbb{1}\{Y_{i,t-1} = q\} + X_{it,q} \beta,$$

with $T = 3$ periods. The target parameter is $\theta_0 = (\beta_0, \gamma_{01}, \dots, \gamma_{0Q}) \in \Theta \subset \mathbb{R}^{Q+1}$ and Proposition 1 shows that it is uniquely identified under support conditions that do not require constancy of the covariates across certain periods. We impose the following assumptions.

Assumption 1 (RECTANGULAR COMPACT PARAMETER SPACE). *The parameter space is*

$$\Theta = [\underline{\beta}, \bar{\beta}] \times [\underline{\gamma}, \bar{\gamma}]^Q$$

for scalars $\underline{\beta} < 0 < \bar{\beta}$, $\underline{\gamma} < 0 < \bar{\gamma}$.

Assumption 2 (IDENTIFICATION RESTRICTIONS). *The outcomes $Y = (Y_1, Y_2, Y_3)$ are generated according to model (7) with $z(\cdot, \cdot, \cdot, \cdot)$ defined in (4) for $Q \geq 3$,¹³ $T = 3$, and true parameter $\theta_0 = (\beta_0, \gamma_{01}, \dots, \gamma_{0Q}) \in \Theta$. Moreover, for any initial condition $y_0 \in \{1, \dots, Q\}$, there exists a non-empty set $\mathcal{X}_{y_0}^{reg} \subseteq \mathbb{R}^{3Q}$ and a compact set $\mathcal{A}_{y_0}^{reg}$ such that for all $x \in \mathcal{X}_{y_0}^{reg}$, $P(A \in \mathcal{A}_{y_0}^{reg} | Y_0 = y_0, X = x)$ is well defined and bounded away from zero.*

Assumption 1 is standard. Assumption 2 imposes the regularity conditions employed in our identification argument. In particular, it guarantees that, the model probabilities of all choice histories $y = (y_1, y_2, y_3)$ remain strictly positive for each initial condition y_0 .

Our identification strategy exploits the conditional moment functions in (19). With $T = 3$ periods, they take the form

$$m_{y_0}^{\sigma_1, \sigma_+, q}(y, x, \theta) \equiv m_{y_0}^{\sigma_1, q}(y, x, \theta) = \mathbb{1}\{y_2 = q\} S_{y_0}^{\sigma_1}(y, x, \theta) - \mathbb{1}\{y_1 = q\}, \quad (37)$$

where σ_+ is the cyclical increment and σ_1 is the reverse cycle. The core idea behind Proposition 1 is to craft a just-identified system of moment equations $\Xi(\theta)$ based on (37) whose

¹³We focus our attention on the case $Q \geq 3$ as Honoré and Weidner (2025) already provided sufficient conditions for identification in the binary case $Q = 2$.

Jacobian $\nabla\Xi(\theta)$ forms a P -matrix on the whole parameter space Θ ; that is, a matrix with all principal minors positive. This property allows us to invoke a global univalence theorem of [Gale and Nikaido \(1965\)](#), which implies that $\Xi(\theta)$ is injective on Θ . Since the moment equations are satisfied at the true parameter, injectivity then yields global identification of θ_0 as the unique root of $\Xi(\theta) = 0$.

The implementation of this idea requires introducing the following covariate regions. Fix a constant $B \in \mathbb{R}_{++}$ with $B > 1/2$. For each $q \in \{1, \dots, Q\}$ and $\epsilon > 0$, let

$$\begin{aligned} \mathcal{X}_{q,\epsilon}^\gamma = \Big\{ & x \in \mathcal{X}_q^{reg} : \|x\|_\infty < B, \\ & x_{1q} \geq x_{3q} + \frac{\epsilon}{3}, \quad x_{3q} \geq x_{2q} + \frac{\epsilon}{3}, \\ & x_{2r} \geq x_{3r} + \frac{\epsilon}{3}, \quad x_{3r} \geq x_{1r} + \frac{\epsilon}{3}, \quad r \in \{1, \dots, Q\} \setminus \{q\}, \\ & |x_{ts,l}| < \epsilon, \quad \forall (t, s, l) \in \{1, 2, 3\}^2 \times \{1, \dots, Q\} \Big\} \end{aligned}$$

and let

$$\begin{aligned} \mathcal{X}_q^\beta = \Big\{ & x \in \mathcal{X}_q^{reg} : \|x\|_\infty < B, \\ & x_{1q} > x_{3q} + \frac{1}{2}, \quad x_{3q} > x_{2q} + \frac{1}{2}, \\ & x_{2r} > x_{3r} + \frac{1}{2}, \quad x_{3r} > x_{1r} + \frac{1}{2}, \quad r \in \{1, \dots, Q\} \setminus \{q\} \Big\}. \end{aligned}$$

where we use the shorthand $x_{ts,q} = x_{t,q} - x_{s,q}$ for conciseness. The set $\mathcal{X}_q^\beta$ can be thought of as identifying the common slope coefficient β , while the sets $\mathcal{X}_{q,\epsilon}^\gamma$ pin down the state-dependence parameters γ_q . These bounded covariate regions are constructed in such a way that the Jacobian of the identifying map, $\nabla\Xi(\theta)$, possesses a particular sign and dominance structure. Specifically, for $\Xi(\theta)$ defined in (40), they imply that $\nabla\Xi(\theta)$ has all entries positive and is strictly diagonally dominant, uniformly over $\theta \in \Theta$, once $\epsilon_1, \dots, \epsilon_Q$ are chosen sufficiently small. The latter condition means that the relevant own-derivative terms dominate the cross-derivative terms in absolute value. Appendix Lemma 6 shows that these properties imply that $\nabla\Xi(\theta)$ is a P -matrix, as desired.

Proposition 1 *Suppose that Assumptions 1-2 hold and that*

$$P(Y_0 = y_0, X \in \mathcal{X}_{y_0}^\beta) > 0 \quad \text{for all } y_0 \in \{1, \dots, Q\} \quad (38)$$

$$P(Y_0 = y_0, X \in \mathcal{X}_{y_0, \epsilon}^\gamma) > 0 \quad \text{for all } y_0 \in \{1, \dots, Q\} \text{ and } \epsilon \in (0, 3B) \quad (39)$$

Then, there exists $\epsilon_1, \dots, \epsilon_Q$, sufficiently small positive constants, such that the mapping

$$\Xi(\theta) = \begin{pmatrix} \mathbb{E} \left[m_{y_0}^{\sigma_1, 1}(Y, X, \theta) | Y_0 = 1, X \in \mathcal{X}_1^\beta \right] \\ \mathbb{E} \left[m_{y_0}^{\sigma_1, 1}(Y, X, \theta) | Y_0 = 1, X \in \mathcal{X}_{1, \epsilon_1}^\gamma \right] \\ \mathbb{E} \left[m_{y_0}^{\sigma_1, 2}(Y, X, \theta) | Y_0 = 2, X \in \mathcal{X}_{2, \epsilon_2}^\gamma \right] \\ \vdots \\ \mathbb{E} \left[m_{y_0}^{\sigma_1, Q}(Y, X, \theta) | Y_0 = Q, X \in \mathcal{X}_{Q, \epsilon_Q}^\gamma \right] \end{pmatrix} \quad (40)$$

is injective on Θ . Thus, θ_0 is the unique solution to $\Xi(\theta) = 0$, i.e., it is identified.

The proof is given in the Appendix. The use of the initial condition $Y_0 = 1$ in the first component of $\Xi(\theta)$, which identifies β , is inessential and could be replaced by any other initial condition.¹⁴ By contrast, our proof recovers the state-dependence parameters by exploiting observations from each initial condition, which explains the support requirement in (39).

In sum, Proposition 1 establishes point identification of the parameter θ_0 in the empirical specification (4) through the moment map $\Xi(\theta)$. Because these moments are evaluated over events with positive probability, the argument also shows that θ_0 is, in principle, $\sqrt{n}$ -estimable. The particular map in (40), however, is best viewed as a constructive device for proving identification. Using it directly for estimation would require selecting the tuning constants $\epsilon_1, \dots, \epsilon_Q$ and restricting attention to small covariate regions, which would be cumbersome and generally inefficient. For this reason, our empirical application instead uses a tuning-parameter-free and more flexible GMM procedure based on the collection of moment functions in (19).

¹⁴Accordingly, (38) can be relaxed to there exists $y_0 \in \{1, \dots, Q\}$ such that $P(Y_0 = y_0, X \in \mathcal{X}_{y_0}^\beta) > 0$, without modification to the proof.

6 Empirical application

In this section, we illustrate the use of the moment conditions derived above for estimation of a simple model for product choice as a function of prices in the presence of state dependence and persistent unobserved heterogeneity. The product that we consider is soft drinks, and the parameters of interest are a common coefficient on price in the indirect utility function and product-specific coefficients on an indicator for whether the same product was chosen on the previous purchase occasion.

The data are from 2019 and come from NielsenIQ.¹⁵ There are 8,100 households making a total of 47,933 purchases of soft drinks. The categories are aggregated into Cola, Root Beer/Ginger Flavor, Citrus Flavor, Soft Drink with Unique Flavor, and Other. The households are located in one of the ten largest Nielsen DMAs. For each of the categories, we calculate the price as a weighted average of sales in all stores of the household Nielsen DMA within the week of purchase. In addition to household ID, option chosen and prices, we also have information on the Nielsen DMA, household income, household size and race.¹⁶ Table 2 displays summary statistics for the data.

Figure 1 shows the histogram of the number of purchases per household and Table 3 shows the number of transitions between products.

Table 3 shows substantial persistence in choices, with 59% of purchases repeating the previous category. However, persistence varies considerably by product. For example, 66% of cola purchases are followed by another cola purchase, while only 49% of citrus drink purchases are followed by citrus drinks. The large degree of persistence suggests that true state dependence and/or persistent unobserved preferences play an important role. These patterns

¹⁵The results presented here are based on our own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researchers and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

¹⁶A Designated Market Area (DMA) is a federally registered trademark and service mark of The Nielsen Company (US), LLC. Income is measured in intervals on the real line with cutoff at \$5,000, \$8,000, \$10,000, \$15,000, \$20,000, \$25,000, \$30,000, \$35,000, \$40,000, \$45,000, \$50,000, \$60,000, and \$100,000. We record income as 2.5 for observations in the bottom interval and as 125 for observations in the top interval; for the remaining observations income is recorded at the midpoint of the interval divided by 1000.

Table 2: Summary Statistics

	No. Obs.	Mean	Std. Dev.
Price (Cola)	47,933	3.190	0.628
Price (Root Beer/Ginger)	47,933	3.261	0.664
Price (Citrus)	47,933	3.308	0.671
Price (Unique Flavor)	47,933	3.348	0.944
Price (Other)	47,933	4.491	1.879
Household size	47,933	2.671	1.384
Income (in \$1,000)	47,933	72.940	37.644
White	47,933	0.715	0.451
Black	47,933	0.173	0.379
Asian	47,933	0.042	0.200
Other	47,933	0.069	0.254

The table displays summary statistics for the variables used here. The data comes from NielsenIQ, and covers 47,933 purchases in the ten largest Nielsen DMAs in 2019. A unit of observation is a purchase occasion. Prices are measured in cents per ounce.

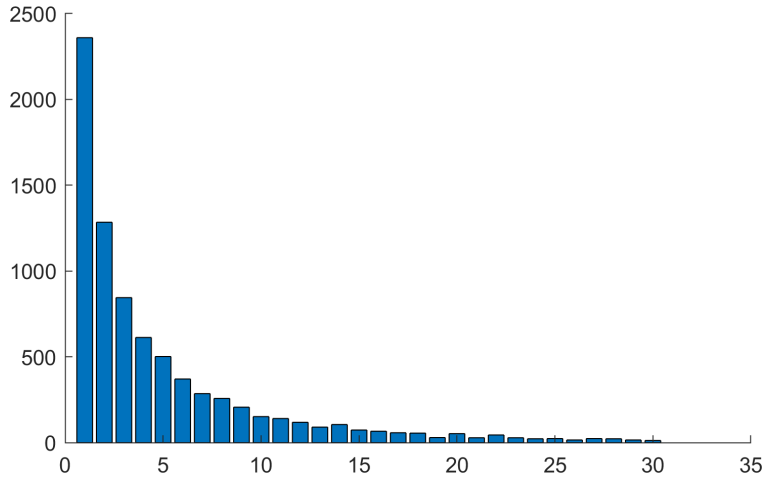
motivate a model that allows for product-specific state dependence as well as unrestricted individual-product heterogeneity. Specifically, we consider the linear specification (4) with price as an explanatory variable. That is,

$$z_t(q, y_{t-1}, \text{Price}, \theta) = \text{Price}_{t,q} \times \beta + \gamma_q \mathbb{1}\{y_{t-1} = q\}$$

where $\theta = (\beta', \gamma_1, \dots, \gamma_Q)'$. In this model, the γ_q 's measure the product specific state dependence and the fixed effects A_{qi} control unrestricted individual-product specific heterogeneity. Proposition 1 in the previous section established that θ is identified under appropriate support conditions.

Let $Z_i^{s:t} = (Z_{is}, Z_{i,s+1}, \dots, Z_{it})$ for $s < t$, and let T_i denote the number of observed purchase occasions for individual i after the initial condition. We estimate the model by generalized method of moments (GMM), using as building blocks the three-period conditional moment functions displayed in (37). For every individual i with $T_i \geq 3$, we exploit the moment functions associated with each consecutive three-period window. That is, for each

Figure 1: Histogram of Purchases per Household



The figure shows the histogram for the number of soft drink purchases for 8,100 households in the ten largest Nielsen DMAs in 2019. 186 households with more than 30 purchases are not displayed. The data comes from NielsenIQ.

Table 3: Transitions Between Products

	Cola	Root Beer/Ginger	Citrus	Unique Flavor	Other
Cola	8528	1142	1179	600	1435
Root Beer/Ginger	1137	3911	639	260	1285
Citrus	1142	620	2871	351	825
Unique Flavor	603	261	360	1836	355
Other	1468	1255	851	371	6548

The table shows the number of times a purchase of a type of soft drink (listed in the column to the left) is followed by a purchase of a possibly different type (listed in the row on top). The data comes from NielsenIQ, and covers 47,933 purchases in the ten largest Nielsen DMAs in 2019.

soda category q , the collection of conditional moment functions we consider is

$$m_{Y_{i0}}^{\sigma_1, \sigma_+, q}(Y_i^{1:3}, \text{Price}_i^{1:3}, \theta), \quad m_{Y_{i1}}^{\sigma_1, \sigma_+, q}(Y_i^{2:4}, \text{Price}_i^{2:4}, \theta), \quad \dots, \quad m_{Y_{i, T_i-3}}^{\sigma_1, \sigma_+, q}(Y_i^{T_i-2:T_i}, \text{Price}_i^{T_i-2:T_i}, \theta).$$

Following [Honoré and Weidner \(2025\)](#), we first pre-normalize each conditional moment function by dividing it by the sum of the absolute values of the distinct summands appearing in its expression.¹⁷ This normalization uniformly bounds the conditional moment functions

¹⁷Appendix equation (59) gives a disaggregated expression of $m_{Y_0}^{\sigma_1, \sigma_+, q}(Y, X, \theta)$, making explicit its constituent terms and the normalization we use.

and their gradients over the parameter space. As in the binary choice setting considered by [Honoré and Weidner \(2025\)](#), we find that this rescaling improves the finite-sample performance of the resulting GMM estimators.

Next, we form unconditional moment functions by interacting the rescaled conditional moment functions $\tilde{m}_{Y_{i,s-1}}^{\sigma_1, \sigma_+, q}(Y_i^{s:s+2}, \text{Price}_i^{s:s+2}, \theta)$ with indicators for the initial condition $Y_{i,s-1}$ and the relative price differences $\text{Price}_{is,q'} - \text{Price}_{is,\text{cola}}$, $\text{Price}_{is+1,q'} - \text{Price}_{is+1,\text{cola}}$, and $\text{Price}_{is+2,q'} - \text{Price}_{is+2,\text{cola}}$ for each product category q' other than cola.¹⁸

Finally, to keep the dimension of the estimating moment vector manageable, we sum up these unconditional moments across all three-period windows within each soda category q , for each individual i . This procedure ultimately yields a total of 85 moment conditions. Since this number is large relative to the sample size,¹⁹ estimating a fully unrestricted optimal weighting matrix may be unstable. We therefore use a two-step GMM procedure in which the second-step weighting matrix is diagonal, with each diagonal element equal to the inverse of the sample variance of the corresponding moment function evaluated at the initial GMM estimator.

The calculations in Section 4.3 above reveal that moment conditions derived from (26) can have heavy-tailed distributions. The results presented in Table 4 suggest that this is not an issue for the moments used in the estimation in this section. Specifically, Table 4 presents summary statistics for the empirical distribution of the 3-period conditional moment functions in equation (37) evaluated at the GMM estimator. Since each of the moments has a high probability of being 0, we present that probability along with summary statistics for the moment functions conditional on not being 0. Conditional on not being 0, all of the five moment functions have skewness close to 0 and the kurtosis is between 1.24 and 1.35. Consistent with equation (37), the minimum value is always -1 , while the maximum value is below 2.5.

In addition to the GMM estimator based on the moments derived in this paper, we also calculate the maximum likelihood estimator of a multinomial logit models that uses prices,

¹⁸Because only price differences matter in the multinomial logit model, we use prices relative to cola rather than price levels.

¹⁹The weighting matrix has 3,655 unique elements, while the sample size is 36,859.

Table 4: Summary Statistics for the Moment Functions

q	$\Pr(m^{\sigma_1, \sigma_+, q} = 0)$	Summary Statistics Conditional on $m^{\sigma_1, \sigma_+, q} \neq 0$							
		Mean	Variance	SD	Skewness	Kurtosis	Min	Median	Max
1	0.705	0.010	0.846	0.920	0.061	1.348	-1.000	-0.003	2.243
2	0.791	-0.009	0.865	0.930	0.073	1.284	-1.000	0.001	2.415
3	0.811	-0.013	0.892	0.944	0.080	1.241	-1.000	-0.010	2.320
4	0.906	0.025	0.947	0.973	0.063	1.243	-1.000	0.001	1.856
5	0.755	-0.014	0.854	0.924	0.088	1.313	-1.000	-0.008	2.235

past choice, household size, household income, and race dummies as explanatory variables. The GMM estimation requires at least four purchase occasions for a household, while the multinomial logit maximum likelihood estimator only requires two. We therefore calculate maximum likelihood estimator on the sample of household with at least two purchases as well as for the data used for the GMM estimator. The results are reported in Table 5.

The maximum likelihood estimates that do not account for unobserved heterogeneity imply substantial state dependence, with coefficients between 1.5 and 2.5 depending on the product category. In contrast, the GMM estimates that control for fixed effects are dramatically smaller, ranging from 0.02 to 0.29. This difference suggests that a large share of the observed persistence in choices is due to persistent unobserved heterogeneity rather than true state dependence. After controlling for this heterogeneity, state dependence remains economically modest and is close to zero for some categories, such as citrus and other soft drinks. The large difference between the MLE and GMM estimates highlights the importance of controlling for unobserved heterogeneity when estimating state dependence in discrete choice models.

In contrast, the estimated price coefficient is remarkably stable across specifications, at approximately -0.11 , suggesting that ignoring unobserved heterogeneity primarily biases estimates of state dependence rather than price sensitivity. At a superficial level, this is not surprising since the prices are plausibly independent of the household fixed effects.

Table 5: Estimation Results: State Dependence and Price Sensitivity

	MLE without heterogeneity		GMM Allowing for Fixed Effects
	≥ 2 purchases	≥ 4 purchases	≥ 4 purchases
Cola (γ)	1.525 (0.050)	1.523 (0.054)	0.140 (0.063)
Root/Ginger (γ)	1.564 (0.065)	1.611 (0.071)	0.164 (0.076)
Citrus (γ)	1.514 (0.070)	1.542 (0.074)	0.017 (0.081)
Unique Flavor (γ)	2.493 (0.101)	2.520 (0.106)	0.293 (0.110)
Other (γ)	1.596 (0.055)	1.644 (0.059)	0.081 (0.068)
Price (β)	-0.110 (0.017)	-0.115 (0.019)	-0.110 (0.025)
Observations	39,833	36,859	36,859

The table shows the results from maximum likelihood estimation and GMM estimation of a dynamic multinomial logit model. The data comes from NielsenIQ and covers 47,933 purchases in the ten largest Nielsen DMAs in 2019. The models estimated by maximum likelihood also include the time invariant variables household income, household size and race dummies for White, Black and Asian as explanatory variables. The GMM specification allows for unrestricted fixed effects. Asymptotic standard errors are reported in parentheses.

7 Conclusion

This paper has shown how to construct moment conditions free of the fixed effects for the dynamic multinomial logit model. The construction rests on a transformed-model device: permuting outcomes across time periods leaves the contribution of the alternative-specific fixed effects unchanged, so that the ratio of a transformed-model probability to the true-model probability is free of those effects. When the permutation is cyclic the transformed model is itself a proper distribution. This alone yields a valid moment condition, and contrasting two such models also delivers moment conditions that depend on the common parameters alone. These conditions require at least three time periods after the initial condition, and we provide an explicit formula for the number of linearly independent moments that are generated this

way. This number grows quickly in both the number of alternatives and the number of time periods.

Two features of the construction of moment conditions in this paper are worth emphasizing. First, the construction is explicit. Unlike the abstract characterization obtained from functional differencing, our approach delivers moment functions in closed form that can be used directly for estimation. Second, the construction is general. The only model-specific ingredient is that the static counterpart of the model admits a sufficient statistic for the fixed effects that is symmetric in the time indices. Whenever such a sufficient statistic exists, the same argument can be used to produce fixed effects-free moment conditions for the corresponding dynamic model with lagged dependent variables. This unifies the multinomial logit with other results, and provides one way to proceed in settings that have not yet been studied.

We put our moment conditions to use in two ways. We establish point identification of the parameters in our empirical specification under support conditions that, unlike conditional-likelihood arguments, do not require covariates to be constant across periods. We also apply the moment conditions to estimate a multinomial logit model with fixed effects in an application to household soft drink purchases. Accounting for unobserved heterogeneity in this application sharply reduces the estimated degree of state dependence relative to a model that ignores it, while leaving the price coefficient essentially unchanged. This suggests that much of the persistence in observed choices can be explained by stable unobserved preferences rather than genuine state dependence.

Several questions remain open. For example, our count of linearly independent moment conditions is supported by a proof for the binary case and otherwise for three time periods. We verified the validity of the claim numerically more broadly, but a general analytical proof for more than two alternatives and more than three periods is still lacking. Also, while the transformed-model moments are always valid, they need not be well behaved when applied in specific settings. As the linear panel data Gaussian example illustrates, they can be heavy-tailed, and they are therefore best viewed as a constructive existence result that complements, rather than replaces, model-specific reasoning about how best to eliminate the fixed effects.

Characterizing efficient or well-behaved combinations of the available moments, and extending the construction to richer forms of dependence and to settings without independence of irrelevant alternatives, are natural directions for future work.

Acknowledgement

This paper benefited from the use of generative AI tools to assist with language editing and L^AT_EX formatting; all output was carefully reviewed by the authors. All substantive content, results, and any remaining errors are the authors' responsibility.

References

- Aguirregabiria, V., J. Gu, and Y. Luo (2021). Sufficient statistics for unobserved heterogeneity in structural dynamic logit models. Journal of Econometrics 223(2), 280–311.
- Al-Sadoon, M. M., T. Li, M. H. Pesaran, O. Yang, and X. Zhao (2026, August). An exponential dynamic multinomial response panel data model with fixed effects. Unpublished manuscript.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. Journal of the Royal Statistical Society: Series B (Methodological) 32(2), 283–301.
- Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Review of Economic Studies 58(2), 277–297.
- Blumenstock, J. E., G. Chi, and X. Tan (2025). Migration and the value of social networks. Review of Economic Studies 92(1), 97–128.
- Bonhomme, S. (2012). Functional differencing. Econometrica 80(4), 1337–1385.
- Chamberlain, G. (1980, January). Analysis of Covariance with Qualitative Data. The Review of Economic Studies 47(1), 225–238.
- Chamberlain, G. (1992). Comment: Sequential moment restrictions in panel data. Journal of Business & Economic Statistics 10(1), 20–26.

- Dano, K. (2023). Transition probabilities and moment restrictions in dynamic fixed effects logit models. arXiv preprint arXiv:2303.00083.
- Davis, D. R., J. I. Dingel, J. Monras, and E. Morales (2019). How segregated is urban consumption? Journal of Political Economy 127(4), 1684–1738.
- Dobronyi, C., J. Gu, K. i. Kim, and T. M. Russell (2021). Identification of dynamic panel logit models with fixed effects. arXiv preprint arXiv:2104.04590.
- Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010). State dependence and alternative explanations for consumer inertia. RAND Journal of Economics 41(3), 417–445.
- Gale, D. and H. Nikaido (1965). The jacobian matrix and global univalence of mappings. Mathematische Annalen 159(2), 81–93.
- Hahn, J. (1997). A note on the efficient semiparametric estimation of some exponential panel models. Econometric Theory 13(4), 583–588.
- Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. American Economic Review 103(7), 2643–2682.
- Heckman, J. J. (1981). Statistical models for discrete panel data. In C. F. Manski and D. McFadden (Eds.), Structural Analysis of Discrete Data with Econometric Applications, pp. 114–178. Cambridge, MA: MIT Press.
- Honoré, B. E. (1993). Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. Journal of Econometrics 59(1–2), 35–61.
- Honoré, B. E. and E. Kyriazidou (2000). Panel data discrete choice models with lagged dependent variables. Econometrica 68(4), 839–874.
- Honoré, B. E. and M. Weidner (2025). Moment conditions for dynamic panel logit models with fixed effects. Review of Economic Studies 92(5), 3112–3137.
- Honoré, B. E., C. Muris, and M. Weidner (2025). Dynamic ordered panel logit models. Quantitative Economics 16(3), 899–945.
- Horn, R. A. and C. R. Johnson (2012). Matrix analysis. Cambridge university press.
- Hu, L. (2002). Estimation of a censored dynamic panel data model. Econometrica 70(6), 2499–2517.

- Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. Journal of Political Economy 105(3), 473–522.
- Khan, S., F. Ouyang, and E. Tamer (2021). Inference on semiparametric multinomial response models. Quantitative Economics 12(3), 743–777.
- Kitazawa, Y. (2001). Exponential regression of dynamic panel data models. Economics Letters 73(1), 7–13.
- Kitazawa, Y. (2022). Transformations and moment conditions for dynamic fixed effects logit models. Journal of Econometrics 229(2), 350–362.
- Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. Review of Economic Studies 68(3), 543–572.
- Lindsay, B. G. (1982). Conditional score functions: some optimality results. Biometrika 69(3), 503–512.
- Magnac, T. (2000). Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories. The economic journal 110(466), 805–837.
- Mathai, A. M. and S. B. Provost (1992). Quadratic Forms in Random Variables: Theory and Applications, Volume 126 of Statistics: Textbooks and Monographs. New York: Marcel Dekker.
- Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. Econometrica 16(1), 1–32.
- Pakes, A., J. R. Porter, M. Shepard, and S. Calder-Wang (2021). Unobserved heterogeneity, state dependence, and health plan choices. NBER Working Paper 29025, National Bureau of Economic Research.
- Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Nielsen & Lydiche.
- Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. Econometric Theory 13(5), 667–678.

A Discussion of Conjecture 1

This appendix discusses two complementary strategies for proving Conjecture 1. The first uses the polynomial structure of the model probabilities in the fixed effect parameter to derive a lower bound on the total number of linearly independent moment conditions, which together with a known matching upper bound pins down the count exactly for $Q = 2$. The second argues directly about linear independence of the $\ell(Q, T)$ moments in (24), which we establish rigorously for $T = 3$ and any $Q \geq 2$.

A.1 Strategy 1: polynomial expansion of the likelihood

Following Dobronyi, Gu, Kim and Russell (2021), define $\eta_q := \exp(\alpha_q)$ for $q = 1, \dots, Q$, and $\pi_t(q, r) := \exp(z_t(q, r, x, \theta))$. Without loss of generality, normalize $\eta_1 = 1$.²⁰ The joint probability (7) factors as

$$p(y \mid y_0, x, \theta, \alpha) = \kappa(\eta) \cdot \tilde{p}(y, \eta), \quad (41)$$

where $\kappa(\eta) = \left[\left(\sum_{r=1}^Q \pi_1(r, y_0) \eta_r \right) \prod_{t=2}^T \prod_{l=1}^Q \left(\sum_{r=1}^Q \pi_t(r, l) \eta_r \right) \right]^{-1}$ depends on η but not on y , and

$$\tilde{p}(y, \eta) = \pi_1(y_1, y_0) \eta_{y_1} \prod_{t=2}^T \prod_{l=1}^Q \left[\sum_{r=1}^Q \pi_t(r, l) \eta_r \right]^{1 - \mathbb{1}\{y_{t-1}=l\}} \pi_t(y_t, y_{t-1}) \eta_{y_t} \quad (42)$$

is a polynomial in $\eta = (\eta_2, \dots, \eta_Q)$ of total degree at most $(T-1)(Q-1) + T = TQ - Q + 1$.

Writing the polynomial in monomial form, $\tilde{p}(y, \eta) = \sum_k c_k(y) \prod_{i=2}^Q \eta_i^{k_i}$, the moment condition (9) (which must hold for all $\eta \in \mathbb{R}_+^{Q-1}$) is equivalent to the system of linear constraints

$$\sum_y m(y) c_k(y) = 0 \quad \text{for every multi-index } k \in \mathbb{N}^{Q-1} \text{ with } |k| \leq TQ - Q + 1. \quad (43)$$

The total number of linearly independent moment conditions is therefore $Q^T - \text{rank}\{c_k(\cdot) : k\}$.

²⁰This is without loss of generality since $p(y \mid y_0, x, \theta, \alpha) = p(y \mid y_0, x, \theta, \alpha^*)$ for $\alpha^* = (0, \alpha_2 - \alpha_1, \dots, \alpha_Q - \alpha_1)$.

Since the number of multi-indices k with $|k| \leq TQ - Q + 1$ in $Q - 1$ variables is $\binom{TQ}{Q-1}$, this rank is at most $\binom{TQ}{Q-1}$, yielding the lower bound

$$(\# \text{ linearly independent moment conditions}) \geq Q^T - \binom{TQ}{Q-1}. \quad (44)$$

The case $Q = 2$. For $Q = 2$, $\tilde{p}(y, \eta_2)$ is a univariate polynomial of degree at most $2T - 1$, so the constraints (43) involve exactly $2T$ coefficient vectors $c_0, c_1, \dots, c_{2T-1} \in \mathbb{R}^{2^T}$. The bound (44) gives at least $2^T - 2T = \ell(2, T)$ linearly independent moment conditions. [Honoré and Weidner \(2025\)](#) construct $\ell(2, T)$ explicit linearly independent moment functions, which gives the same lower bound. The matching upper bound is established in subsequent work; see [Dobronyi, Gu, Kim and Russell \(2021\)](#) and [Dano \(2023\)](#). This proves Conjecture 1 for $Q = 2$ and any $T \geq 3$.

The case $Q \geq 3$. For $Q \geq 3$, the bound (44) is not sharp and can even be negative. For example, $\binom{9}{2} = 36 > 27 = 3^3$, so (44) is uninformative at $(Q, T) = (3, 3)$. The reason is that the $\binom{TQ}{Q-1}$ coefficient vectors $c_k(\cdot) \in \mathbb{R}^{Q^T}$ are not all linearly independent: the product structure of $\tilde{p}(y, \eta)$ as a product of linear forms in η forces algebraic relations among the c_k . Identifying these relations in closed form, and thereby sharpening (44), is an open problem.

A.2 Strategy 2: linear independence of the moments in (24)

The second strategy works directly with the $\ell(Q, T)$ explicit moment functions in (24). If these moments are linearly independent for generic (y_0, x, θ) , then at least $\ell(Q, T)$ linearly independent moment conditions exist. Combined with a matching upper bound on that number, which Strategy 1 delivers for $Q = 2$, this establishes Conjecture 1.

We prove linear independence rigorously for $T = 3$ and any $Q \geq 2$. For $T = 3$ the only allowed value is $j = 0$, which forces $\mathbf{t} = \emptyset$, $s_1 = 1$, $s_2 = 2$, and varies $q_0 \in \{1, \dots, Q\}$. The two cyclic permutations of $\{1, 2, 3\}$ have generation orders $\tau_1 = (1, 2, 3)$ (so that σ_1 is the cyclic increment σ_+ and $S^{\sigma_1} = 1$) and $\tau_2 = (2, 1, 3)$ (the reverse cyclic). The parameterization (24) assigns s_1 to σ_1 , so the two labels are interchanged here relative to the $T = 3$ example

of Section 3.3, where σ_1 denotes the reverse cycle. The resulting Q moment functions are

$$m_q(y) := \mathbb{1}\{y_1 = q\} - \mathbb{1}\{y_2 = q\} S_{y_0}^{\sigma_2}(y, x, \theta), \quad q = 1, \dots, Q. \quad (45)$$

Proposition 2 *For $T = 3$ and any $Q \geq 2$, the moment functions $m_1, \dots, m_Q$ in (45) are linearly independent in $\mathbb{R}^{Q^T}$ for generic (y_0, x, θ_0) .*

Proof. For each $q \in \{1, \dots, Q\}$, fix any $c_q \neq q$ and define the evaluation point

$$y_q := (q, q, c_q) \in \{1, \dots, Q\}^3.$$

At y_q , both indicators in (45) reduce to $\mathbb{1}\{q' = q\}$, so for each $q' \in \{1, \dots, Q\}$,

$$m_{q'}(y_q) = \mathbb{1}\{q = q'\} [1 - S_{y_0}^{\sigma_2}(y_q, x, \theta)].$$

The $Q \times Q$ matrix $[m_{q'}(y_q)]_{q', q}$ is therefore diagonal with diagonal entries $1 - S_{y_0}^{\sigma_2}(y_q, x, \theta)$.

It remains to verify that $S_{y_0}^{\sigma_2}(y_q, x, \theta) \neq 1$ for generic (y_0, x, θ_0) . The reverse cyclic σ_2 acts as $\sigma_2(1) = 3, \sigma_2(2) = 1, \sigma_2(3) = 2$, so $\sigma_2(0) := \sigma_2(3) = 2$. Using (16),

$$\begin{aligned} S_{y_0}^{\sigma_2}((q, q, c_q), x, \theta) &= \exp \left[(z_1(q, y_0) - z_1(q, y_0)) + (z_2(c_q, q) - z_2(q, q)) \right. \\ &\quad \left. + (z_3(q, q) - z_3(c_q, q)) \right] \\ &= \exp \left[(z_2(c_q, q) - z_3(c_q, q)) - (z_2(q, q) - z_3(q, q)) \right]. \end{aligned}$$

For the linear specification (3), $z_t(r, l, x, \theta) = \gamma_{r,l} + x'_t \beta_r$, so the γ terms cancel and

$$S_{y_0}^{\sigma_2}(y_q, x, \theta) = \exp[(x_2 - x_3)'(\beta_{c_q} - \beta_q)].$$

For generic x with $x_2 \neq x_3$ and generic β with $\beta_{c_q} \neq \beta_q$ for each q , the exponent is non-zero and $S_{y_0}^{\sigma_2}(y_q, x, \theta) \neq 1$. The diagonal matrix $[m_{q'}(y_q)]$ is therefore invertible, and the moment functions $m_1, \dots, m_Q$ are linearly independent. ■

Proposition 2 establishes the first half of Conjecture 1 (linear independence of the con-

structed moments) for $T = 3$ and any $Q \geq 2$, and implies that at least $\ell(Q, 3) = Q$ linearly independent moment conditions of the form (8) exist for generic (y_0, x, θ_0) . The second half of the conjecture (that every valid moment function is a linear combination of the constructed moments, equivalently that there are no more than $\ell(Q, 3)$ linearly independent moment conditions) is supported by direct numerical computation for $T = 3$ but does not have a clean analytical proof for $Q \geq 3$.

The obstacle for $T \geq 4$. For $T \geq 4$, linear independence of the $\ell(Q, T)$ moment functions in (24) has been verified by direct numerical computation for all values of (Q, T) in Table 1, but a general analytical proof has so far eluded us. A natural extension of the $T = 3$ argument groups the moments hierarchically by the number of shared time indices j and, within each level, by the shared-position values $\mathbf{t}$ and $\mathbf{q}$. Within a fixed group the moments reduce to a structure analogous to the $T = 3$ case, and a similar diagonal evaluation-point argument applies on a restricted slice of the outcome space. Combining the slices into a single linear-independence statement requires bookkeeping that we have not been able to carry out cleanly for arbitrary (Q, T) .

A.3 Summary

Combining the two strategies, Conjecture 1 is established analytically for $Q = 2$ and any $T \geq 3$ (by Strategy 1) and, as a lower bound, for $T = 3$ and any $Q \geq 2$ (by Strategy 2). For $Q \geq 3$ the conjecture remains open in both halves: at $T = 3$ we have the lower bound but not exhaustiveness, and for $T \geq 4$ neither half is established. Equality has been verified numerically for all (Q, T) in Table 1 but not proved analytically. We leave a complete analytical proof of Conjecture 1 for $Q \geq 3$ for future work.

B Generalized moment conditions for non-cyclic permutations

Theorem 1 constructs moment conditions from pairs of cyclic permutations σ_1, σ_2 . In this appendix, we show that the construction extends to pairs of general (not necessarily cyclic) permutations, provided they satisfy a compatibility condition on their cycle structures. In all cases we have checked numerically, the resulting moment conditions do not generate any linearly independent moments beyond those already obtained from cyclic permutations in Theorem 1, so this generalization is not needed for Conjecture 1. However, the result is of independent interest, as it clarifies the structure of the transformed model approach and may be useful for extensions to other models.

B.1 Transformed models for non-cyclic permutations

For a general (not necessarily cyclic) permutation $\sigma : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$, the transformed model p^σ defined in (11) does not correspond to a single Markov chain through all T outcomes. Instead, p^σ factorizes as a product of independent Markov chains, one for each cycle of σ .

To see this, recall that every permutation σ can be decomposed into disjoint cycles. Let $\sigma = c_1 \circ c_2 \circ \dots \circ c_L$ be the cycle decomposition, with $L \geq 1$ cycles. Each cycle c_ℓ acts on a subset $C_\ell \subset \{1, \dots, T\}$, and $\{C_1, \dots, C_L\}$ is a partition of $\{1, \dots, T\}$. Let C_* denote the cycle containing T , and write $C_* = \{T, \sigma(T), \sigma^2(T), \dots, \sigma^{|C_*|-1}(T)\}$. For the cycle C_* , define the generation order $\tau_*(k) := \sigma^k(T)$ for $k = 1, \dots, |C_*|$, so that $\tau_*(|C_*|) = T$. Similarly, for each cycle C_ℓ not containing T , pick any element $a_\ell \in C_\ell$ and define the generation order within that cycle by $\tau_\ell(k) := \sigma^k(a_\ell)$ for $k = 1, \dots, |C_\ell|$, with $\tau_\ell(|C_\ell|) = a_\ell$. In (46) below we also use the convention $\tau_\ell(0) := a_\ell$, which reflects that this chain returns to its starting outcome.

The transformed model p^σ can then be written as

$$\begin{aligned}
p^\sigma(y \mid y_0, x, \theta, \alpha) &= p_1(y_{\tau_*(1)} \mid y_0, x, \theta, \alpha) \underbrace{\prod_{k=2}^{|C_*|} p_{\tau_*(k-1)+1}(y_{\tau_*(k)} \mid y_{\tau_*(k-1)}, x, \theta, \alpha)}_{\text{chain through cycle } C_*} \\
&\times \prod_{\ell: C_\ell \neq C_*} \underbrace{\prod_{k=1}^{|C_\ell|} p_{\tau_\ell(k-1)+1}(y_{\tau_\ell(k)} \mid y_{\tau_\ell(k-1)}, x, \theta, \alpha)}_{\text{chain through cycle } C_\ell}.
\end{aligned} \tag{46}$$

Each factor in (46) is a Markov chain that generates the outcomes $\{y_s : s \in C_\ell\}$ independently of the outcomes in other cycles. For cycles $C_\ell \neq C_*$, the chain starts from the outcome y_{a_ℓ} , which also appears as the outcome generated in the last step of the chain (since $\tau_\ell(|C_\ell|) = a_\ell$). Hence, summing p^σ over all y_s with $s \in C_\ell$ for $C_\ell \neq C_*$ does not yield one for generic (x, θ, α) , because these chains are “closed loops” that both start and end at y_{a_ℓ} .

Nevertheless, Lemma 2 continues to hold: the ratio $S_{y_0}^\sigma(y, x, \theta) = p^\sigma(y \mid y_0, x, \theta, \alpha)/p(y \mid y_0, x, \theta, \alpha)$ is free of α for any permutation σ , not just cyclic ones. This is because the proof of Lemma 2 only uses the fact that σ is a permutation (so that $\sum_t \alpha_{y_{\sigma(t-1)}} = \sum_t \alpha_{y_t}$), not that σ is cyclic.

B.2 Equivalence relation on permutations

We define an equivalence relation $\sim$ on the set of all permutations of $\{1, \dots, T\}$ as follows.

Definition 1 *Two permutations $\sigma_1, \sigma_2 : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ satisfy $\sigma_1 \sim \sigma_2$ if and only if their cycle decompositions are identical except possibly for the cycles containing T .*

That is, $\sigma_1 \sim \sigma_2$ if every cycle that does not contain T appears identically in both σ_1 and σ_2 . The cycles containing T are allowed to differ but must have the same length. Note that $\sigma \sim \sigma$ always holds.

For example, with $T = 7$, the permutations $\sigma_1 = [1, 5, 3][2, 4, 7][6]$ and $\sigma_2 = [1, 5, 3][2, 7, 4][6]$ satisfy $\sigma_1 \sim \sigma_2$, because the cycles $[1, 5, 3]$ and $[6]$ (which do not contain $T = 7$) are identical, while the cycles $[2, 4, 7]$ and $[2, 7, 4]$ (which contain $T = 7$) differ.

When $\sigma_1 \sim \sigma_2$, the two permutations share the same “non- T cycles.” Denote their union by $C_{\text{shared}} := \bigcup_{\ell: T \notin C_\ell} C_\ell$. Let $C_*^{(1)}$ and $C_*^{(2)}$ denote the cycles of σ_1 and σ_2 that contain T , respectively. Since $\sigma_1 \sim \sigma_2$, we have $C_*^{(1)} = C_*^{(2)}$ as sets (both equal $\{1, \dots, T\} \setminus C_{\text{shared}}$), though σ_1 and σ_2 may permute the elements of this set differently.

B.3 Shared index set

For two permutations $\sigma_1 \sim \sigma_2$, we now define the set of “shared indices” that determines the structure of the generalized moment conditions. Consider the generation orders within the cycles containing T :

$$\tau_*^{(1)}(k) := \sigma_1^k(T), \quad \tau_*^{(2)}(k) := \sigma_2^k(T), \quad k = 1, 2, \dots, |C_*|.$$

Since the two cycles contain the same set of elements $C_*^{(1)} = C_*^{(2)}$, both generation orders are permutations of the same set, but they may list the elements in different orders. Additionally, the elements of C_{shared} provide further shared indices from the non- T cycles.

We define the shared index set as

$$\mathcal{I}_{\sigma_1, \sigma_2} := \{(t, t) : t \in C_{\text{shared}}\} \cup \{(t, t) : t = \sigma_1^j(T) = \sigma_2^j(T), j = 1, \dots, k^*\} \cup \{(s_1, s_2)\}, \quad (47)$$

where the diagonal pairs (t_k, t_k) correspond to all elements in the shared non- T cycles together with the elements on which the two generation orders through C_* still agree, and (s_1, s_2) with $s_1 \neq s_2$ corresponds to the “split point” where the two cycles containing T first diverge. More precisely, let k^* be the largest integer such that $\tau_*^{(1)}$ and $\tau_*^{(2)}$ agree on their first k^* entries:

$$\sigma_1^j(T) = \sigma_2^j(T), \quad j = 1, \dots, k^*,$$

and $\sigma_1^{k^*+1}(T) \neq \sigma_2^{k^*+1}(T)$. Then $s_1 = \sigma_1^{k^*+1}(T)$ and $s_2 = \sigma_2^{k^*+1}(T)$, and the t_k from the T -cycle are the shared elements $\sigma_1^j(T) = \sigma_2^j(T)$ for $j = 1, \dots, k^*$. Both generation orders start from y_0 , so it is the shared initial segment of $\tau_*^{(1)}$ and $\tau_*^{(2)}$ that matters here, exactly as

in Theorem 1.

We write $[\sigma_1, \sigma_2] := |\mathcal{I}_{\sigma_1, \sigma_2}|$ for the total number of shared index pairs.

B.4 Generalized theorem

Theorem 2 *Let $\sigma_1, \sigma_2 : \{1, \dots, T\} \rightarrow \{1, \dots, T\}$ be two permutations satisfying $\sigma_1 \sim \sigma_2$ (Definition 1), with $\sigma_1 \neq \sigma_2$. Let $\mathcal{I}_{\sigma_1, \sigma_2} = \{(t_k, u_k) : k = 1, \dots, [\sigma_1, \sigma_2]\}$ with $t_1 < t_2 < \dots < t_{[\sigma_1, \sigma_2]}$ be as in (47). For any $q \in \{1, \dots, Q\}^{[\sigma_1, \sigma_2]}$, define the moment function*

$$m_{y_0}^{\sigma_1, \sigma_2, q}(y, x, \theta) := \mathbb{1}\left\{(y_{t_1}, \dots, y_{t_{[\sigma_1, \sigma_2]}}) = q\right\} S_{y_0}^{\sigma_1}(y, x, \theta) - \mathbb{1}\left\{(y_{u_1}, \dots, y_{u_{[\sigma_1, \sigma_2]}}) = q\right\} S_{y_0}^{\sigma_2}(y, x, \theta). \quad (48)$$

Then

$$\mathbb{E}\left[m_{Y_0}^{\sigma_1, \sigma_2, q}(Y, X, \theta_0) \mid Y_0, X, A\right] = 0. \quad (49)$$

When both σ_1 and σ_2 are cyclic permutations, the shared index set $\mathcal{I}_{\sigma_1, \sigma_2}$ reduces to the structure in Theorem 1: the diagonal pairs (t_k, t_k) correspond to the shared initial segment of the generation orders τ_1 and τ_2 , and $(s_1, s_2) = (\tau_1(k), \tau_2(k))$ is the first point of disagreement.

Remark 1 *Analogous to (20) in Theorem 1, the result can be written equivalently as: for any $\tilde{q} \in \{1, \dots, Q\}$,*

$$\mathbb{E}\left[\mathbb{1}\{Y_{s_1} = \tilde{q}\} S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - \mathbb{1}\{Y_{s_2} = \tilde{q}\} S_{Y_0}^{\sigma_2}(Y, X, \theta_0) \mid \{Y_{t_k}\}_{k=1}^{[\sigma_1, \sigma_2]-1}, Y_0, X, A\right] = 0,$$

where $\{(t_k, t_k) : k = 1, \dots, [\sigma_1, \sigma_2] - 1\}$ are the diagonal elements of $\mathcal{I}_{\sigma_1, \sigma_2}$ and (s_1, s_2) is the off-diagonal element.

B.5 Proof of Theorem 2

By Lemma 2 (which holds for all permutations, not just cyclic ones), we have $S_{y_0}^{\sigma_i}(y, x, \theta) \cdot p(y \mid y_0, x, \theta, \alpha) = p^{\sigma_i}(y \mid y_0, x, \theta, \alpha)$ for $i \in \{1, 2\}$. Therefore, establishing (49) is equivalent to

showing that for all α ,

$$\sum_{y \in \{1, \dots, Q\}^T} \left[\mathbb{1}\left\{(y_{t_1}, \dots, y_{t_{[\sigma_1, \sigma_2]}}) = q\right\} p^{\sigma_1}(y) - \mathbb{1}\left\{(y_{u_1}, \dots, y_{u_{[\sigma_1, \sigma_2]}}) = q\right\} p^{\sigma_2}(y) \right] = 0, \quad (50)$$

where we suppress the dependence on (y_0, x, θ, α) for brevity.

Step 1: Factorization of the transformed model over cycles. For any permutation σ with cycle decomposition $\sigma = c_1 \circ \dots \circ c_L$, the transformed model (15) can be written as a product over cycles:

$$p^\sigma(y) = \prod_{\ell=1}^L \prod_{\substack{t \in \{1, \dots, T\}: \\ \sigma(t-1) \in C_\ell}} p_t(y_{\sigma(t-1)} \mid y_{t-1}). \quad (51)$$

To see this, note that $p^\sigma(y) = \prod_{t=1}^T p_t(y_{\sigma(t-1)} \mid y_{t-1})$ by (15), and the mapping $t \mapsto \sigma(t-1)$ assigns each factor to the cycle containing $\sigma(t-1)$, giving a partition of the product over cycles. We denote the factor corresponding to cycle C_ℓ by

$$p_{C_\ell}^\sigma(y) := \prod_{\substack{t \in \{1, \dots, T\}: \\ \sigma(t-1) \in C_\ell}} p_t(y_{\sigma(t-1)} \mid y_{t-1}).$$

For the cycle C_* containing T , this factor includes the term at $t = 1$ (since $\sigma(0) = \sigma(T) \in C_*$). Note that $p_{C_\ell}^\sigma(y)$ depends on y only through $\{y_s : s \in C_\ell\}$ and $\{y_{t-1} : \sigma(t-1) \in C_\ell\}$. We have $p^\sigma(y) = p_{C_*}^\sigma(y) \cdot \prod_{\ell: C_\ell \neq C_*} p_{C_\ell}^\sigma(y)$.

Step 2: Shared cycles contribute identically. Since $\sigma_1 \sim \sigma_2$, the non- T cycles are identical. For any shared cycle $C_\ell \neq C_*$, the set $\{t : \sigma_1(t-1) \in C_\ell\}$ and $\{t : \sigma_2(t-1) \in C_\ell\}$ are the same (since σ_1 and σ_2 act identically on C_ℓ). Hence $p_{C_\ell}^{\sigma_1}(y) = p_{C_\ell}^{\sigma_2}(y)$ for all y .

Denote the common product of shared-cycle factors by

$$F_{\text{shared}}(y) := \prod_{\ell: C_\ell \neq C_*} p_{C_\ell}^{\sigma_1}(y) = \prod_{\ell: C_\ell \neq C_*} p_{C_\ell}^{\sigma_2}(y).$$

Then $p^{\sigma_i}(y) = F_{\text{shared}}(y) \cdot p_{C_*}^{\sigma_i}(y)$ for $i = 1, 2$.

Step 3: Reduction to the T -cycle. Write $q = (\bar{q}, \tilde{q})$, where $\bar{q} = (q_1, \dots, q_{[\sigma_1, \sigma_2]-1})$ corresponds to the diagonal elements of $\mathcal{I}_{\sigma_1, \sigma_2}$ and $\tilde{q} \in \{1, \dots, Q\}$ corresponds to the off-diagonal element (s_1, s_2) . Then the left-hand side of (50) becomes

$$\sum_y \mathbb{1}\{y_{t_k} = q_k, k = 1, \dots, [\sigma_1, \sigma_2] - 1\} F_{\text{shared}}(y) [\mathbb{1}\{y_{s_1} = \tilde{q}\} p_{C_*}^{\sigma_1}(y) - \mathbb{1}\{y_{s_2} = \tilde{q}\} p_{C_*}^{\sigma_2}(y)].$$

Now we perform the summation. The indicator $\mathbb{1}\{y_{t_k} = q_k\}$ fixes the outcomes at positions in C_{shared} and on the shared initial segment of C_* . Since $F_{\text{shared}}(y)$ depends on y only through $\{y_s : s \in C_{\text{shared}}\}$ and certain conditioning variables y_{t-1} , and since $p_{C_*}^{\sigma_i}(y)$ depends on y only through $\{y_s : s \in C_*\}$ and certain y_{t-1} , we can separate the summation.

The summation over $\{y_s : s \in C_{\text{shared}}\}$ is constrained by the indicators, and yields F_{shared} evaluated at the fixed values. After this, what remains is a sum over $\{y_s : s \in C_*\}$, with the outcomes on the shared initial segment of C_* also fixed by the indicators:

$$F_{\text{shared}}(\bar{q}) \cdot \sum_{\{y_s : s \in C_*\}} \mathbb{1}\{y_{\sigma_1^j(T)} = \bar{q}_j, j = 1, \dots, k^*\} [\mathbb{1}\{y_{s_1} = \tilde{q}\} p_{C_*}^{\sigma_1}(y) - \mathbb{1}\{y_{s_2} = \tilde{q}\} p_{C_*}^{\sigma_2}(y)],$$

where $F_{\text{shared}}(\bar{q})$ absorbs the shared-cycle contributions.

Step 4: Telescoping within the T -cycle. It remains to show that for each $i \in \{1, 2\}$,

$$H_i := \sum_{\{y_s : s \in C_*\}} \mathbb{1}\{y_{\sigma_i^j(T)} = \bar{q}_j, j = 1, \dots, k^*\} \mathbb{1}\{y_{s_i} = \tilde{q}\} p_{C_*}^{\sigma_i}(y)$$

is the same for $i = 1$ and $i = 2$. Within the T -cycle, $p_{C_*}^{\sigma_i}(y)$ is a product of transition probabilities along the generation order $\tau_*^{(i)}$. Summing over all y_s for $s \in C_*$ that are not fixed by the indicators (i.e., not in the shared initial segment $\{\sigma_i(T), \dots, \sigma_i^{k^*}(T)\}$ and not the split point s_i), the Markov property causes these sums to telescope to one, since $\sum_{r=1}^Q p_t(r \mid \cdot) = 1$.

After the telescoping, we are left with

$$H_i = p_1(\bar{q}_1 \mid y_0) \prod_{j=2}^{k^*} p_{\sigma_i^{j-1}(T)+1}(\bar{q}_j \mid \bar{q}_{j-1}) \cdot p_{\sigma_i^{k^*}(T)+1}(\bar{q} \mid \bar{q}_{k^*}),$$

where $\bar{q}_j$ is the value assigned to the shared index $\sigma_1^j(T) = \sigma_2^j(T)$, and where for $k^* = 0$ the expression reduces to $p_1(\bar{q} \mid y_0)$. Since $\sigma_1^j(T) = \sigma_2^j(T)$ for $j = 0, 1, \dots, k^*$, every factor is the same for $i = 1$ and $i = 2$, so $H_1 = H_2$.

Therefore, the expression in (50) equals $F_{\text{shared}}(\bar{q}) \cdot (H_1 - H_2) = 0$, establishing (49). $\square$

Example: $T = 4$ with non-cyclic permutations. For $T = 4$, consider $\sigma_1 = [1, 2, 4][3]$ and $\sigma_2 = [1, 4, 2][3]$ (in sequence form, $\sigma_1 = (2, 4, 3, 1)$ and $\sigma_2 = (4, 1, 3, 2)$). Both share the cycle $[3]$ (which does not contain $T = 4$), and differ in the cycle containing 4: $[1, 2, 4]$ versus $[1, 4, 2]$. Since neither σ_1 nor σ_2 is a cyclic permutation of $\{1, 2, 3, 4\}$, the moment conditions from Theorem 2 cannot be obtained from Theorem 1. The shared index set is $\mathcal{I}_{\sigma_1, \sigma_2} = \{(3, 3), (1, 2)\}$, with $[\sigma_1, \sigma_2] = 2$, and the corresponding moment conditions are

$$\mathbb{E} \left[\mathbb{1}\{Y_1 = q\} S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - \mathbb{1}\{Y_2 = q\} S_{Y_0}^{\sigma_2}(Y, X, \theta_0) \mid Y_3, Y_0, X, A \right] = 0, \quad (52)$$

for $q \in \{1, \dots, Q\}$. We now verify (52) directly using the chain representation.

The cycle $C_* = \{1, 2, 4\}$ contains $T = 4$, and the shared cycle is $\{3\}$. For σ_1 , the generation order within C_* is $\tau_*^{(1)} = (1, 2, 4)$, meaning the chain visits y_1 , then y_2 , then y_4 . For σ_2 , the generation order is $\tau_*^{(2)} = (2, 1, 4)$, visiting y_2 , then y_1 , then y_4 . The shared cycle $\{3\}$ is a fixed point, producing a closed-loop factor. Using the chain representation from Section B.1, the transformed models are

$$\begin{aligned} p^{\sigma_1}(y) &= \underbrace{p_1(y_1 \mid y_0) \cdot p_2(y_2 \mid y_1) \cdot p_3(y_4 \mid y_2)}_{\text{chain through } C_* = \{1, 2, 4\}} \cdot \underbrace{p_4(y_3 \mid y_3)}_{\text{loop at } \{3\}}, \\ p^{\sigma_2}(y) &= \underbrace{p_1(y_2 \mid y_0) \cdot p_3(y_1 \mid y_2) \cdot p_2(y_4 \mid y_1)}_{\text{chain through } C_* = \{1, 2, 4\}} \cdot \underbrace{p_4(y_3 \mid y_3)}_{\text{loop at } \{3\}}. \end{aligned}$$

The closed-loop factor $p_4(y_3 \mid y_3)$ is identical in both expressions, as guaranteed by $\sigma_1 \sim \sigma_2$.

Using $S^{\sigma_i} \cdot p = p^{\sigma_i}$, the moment condition (52) is equivalent to showing that for all α and for fixed y_3 ,

$$\sum_{y_1, y_2, y_4} \mathbb{1}\{y_1 = q\} p^{\sigma_1}(y) = \sum_{y_1, y_2, y_4} \mathbb{1}\{y_2 = q\} p^{\sigma_2}(y). \quad (53)$$

The common factor $p_4(y_3 \mid y_3)$ cancels from both sides. For the left-hand side, we sum the C_* -chain of σ_1 over the free variables. Sum over y_4 first (the last outcome generated by the chain):

$$\sum_{y_4} p_3(y_4 \mid y_2) = 1 \quad (\text{telescopes}).$$

Then sum over y_2 :

$$\sum_{y_2} p_2(y_2 \mid y_1) = 1 \quad (\text{telescopes}).$$

What remains is $\mathbb{1}\{y_1 = q\} \cdot p_1(y_1 \mid y_0) = p_1(q \mid y_0)$.

For the right-hand side, we sum the C_* -chain of σ_2 analogously. Sum over y_4 :

$$\sum_{y_4} p_2(y_4 \mid y_1) = 1 \quad (\text{telescopes}).$$

Then sum over y_1 :

$$\sum_{y_1} p_3(y_1 \mid y_2) = 1 \quad (\text{telescopes}).$$

What remains is $\mathbb{1}\{y_2 = q\} \cdot p_1(y_2 \mid y_0) = p_1(q \mid y_0)$.

Both sides equal $p_1(q \mid y_0)$, confirming (53). The key mechanism is that both chains start from y_0 (via the p_1 factor) and visit the same set of outcomes $\{y_1, y_2, y_4\}$, but in different orders. When summing over all the free variables, the intermediate transitions telescope to one, and only the initial transition probability $p_1(q \mid y_0)$ survives — regardless of the order in which the chain visits the elements.

C Details for the general exponential–family, bivariate gamma, and linear–normal examples

Lemma 4 states the sufficient statistic for the exponential–family model (28) discussed in Section 4, and Proposition 3 verifies the validity of the moment condition obtained from Lemma 3.

Lemma 4 (SUFFICIENT STATISTIC FOR THE EXPONENTIAL–FAMILY MODEL). *Let $(Y_1, \dots, Y_T)$ be independent with densities $f(y_t, \eta_t, \alpha)$ given by (28), for indices $\eta_1, \dots, \eta_T$ and common fixed effect α . Then $S(Y) = \sum_{t=1}^T s(Y_t)$ is sufficient for α , and the conditional distribution of $(Y_1, \dots, Y_T)$ given $S = \bar{s}$ does not depend on α :*

$$f(y \mid S = \bar{s}, x; \theta) = \frac{\exp\left(\sum_{t=1}^T b(y_t, \eta_t, \theta)\right)}{\int_{\mathcal{A}_{\bar{s}}} \exp\left(\sum_{t=1}^T b(u_t, \eta_t, \theta)\right) du}, \quad \mathcal{A}_{\bar{s}} = \left\{u : \sum_t s(u_t) = \bar{s}\right\}. \quad (54)$$

Moreover S is symmetric in its arguments.

Proof. By (28), the joint density factors as

$$\prod_{t=1}^T f(y_t, \eta_t, \alpha) = \exp(\alpha' e(\theta) S(y)) \cdot \exp\left(\sum_{t=1}^T d(\eta_t, \alpha, \theta)\right) \cdot \exp\left(\sum_{t=1}^T b(y_t, \eta_t, \theta)\right),$$

with $S(y) = \sum_t s(y_t)$. The first factor depends on y only through $S(y)$, and the second is free of y . Conditioning on $S = \bar{s}$ divides by the integral of the joint over $\mathcal{A}_{\bar{s}}$; the factor $\exp(\alpha' e(\theta) \bar{s})$, which is constant on $\mathcal{A}_{\bar{s}}$, and the y –free factor $\exp(\sum_t d(\eta_t, \alpha, \theta))$ cancel between numerator and denominator, leaving (54), which is free of α . Finally, $S = \sum_t s(Y_t)$ is a sum and is therefore symmetric in $(Y_1, \dots, Y_T)$. ■

Proposition 3 (MOMENT CONDITIONS FOR THE EXPONENTIAL–FAMILY PANEL MODEL). *Consider the dynamic model (27) with period densities (28) and $T = 3$. Then for any bounded g ,*

$$\mathbb{E}\left[g(Y_2) S_{Y_0}^{\sigma_1}(Y, X, \theta_0) - g(Y_1) \mid Y_0, X, A\right] = 0, \quad (55)$$

where $S_{Y_0}^{\sigma_1}$ is given by (29) at $\theta = \theta_0$.

Proof. By Lemma 3, $S_{Y_0}^{\sigma_1} p = p^{\sigma_1}$, so

$$\mathbb{E}[g(Y_2) S_{Y_0}^{\sigma_1} \mid Y_0, X, A] = \int g(y_2) p^{\sigma_1}(y \mid Y_0, X, \theta_0, A) dy,$$

where the integral is over the support of (Y_1, Y_2, Y_3) . For the reverse-cyclic permutation,

$$p^{\sigma_1}(y \mid Y_0, X, \theta_0, A) = f(y_2, \eta_1, A) f(y_3, \eta_2, A) f(y_1, \eta_3, A),$$

with $\eta_1 = z_1(Y_0, X, \theta_0)$ constant given Y_0 , $\eta_2 = z_2(y_1, X, \theta_0)$ depending on y_1 but not on y_3 , and $\eta_3 = z_3(y_2, X, \theta_0)$ depending on y_2 but not on y_1 . Integrate sequentially. The factor $f(y_3, \eta_2, A)$ is a density in y_3 , so $\int f(y_3, \eta_2, A) dy_3 = 1$; after eliminating y_3 , the factor $f(y_1, \eta_3, A)$ is a density in y_1 , so $\int f(y_1, \eta_3, A) dy_1 = 1$. The remaining one-dimensional integral is

$$\int g(y_2) f(y_2, \eta_1, A) dy_2 = \mathbb{E}[g(Y_1) \mid Y_0, X, A],$$

because under the true model $Y_1 \mid Y_0, X, A$ has density $f(\cdot, \eta_1, A)$ with the same $\eta_1 = z_1(Y_0, X, \theta_0)$. Combining the three steps gives (55). ■

Remark 2 *The argument is structurally identical to the proof of Lemma 1 for the discrete case. The Markov property ensures that η_t depends on y_{t-1} but not on y_t . As a result, the reverse-cyclic permutation places each Y_t as a free integration variable while its density parameter η_t is governed by a different element of the triple, allowing each factor to integrate to unity independently. Summation in the discrete case is replaced by integration, and the indicator $\mathbf{1}\{Y_{\tau(k)} = q\}$ is replaced by the weighting function g evaluated at $Y_{\tau(k)}$.*

The bivariate gamma model (31) illustrates a model whose symmetric sufficient statistic depends on the common parameter, and therefore lies outside Hahn's class. We record the corresponding sufficient statistic and moment condition.

Lemma 5 (SUFFICIENT STATISTIC FOR THE BIVARIATE GAMMA MODEL). *Let $\{(Y_{1t}, Y_{2t})\}_{t=1}^T$ be independent across t , with $Y_{1t} \sim \text{Gamma}(z_{1t}, A)$ and $Y_{2t} \sim \text{Gamma}(z_{2t}, \phi A)$ for given in-*

dices z_{1t}, z_{2t} in the (shape, rate) parametrization. Then $s_\theta(Y) = \sum_{t=1}^T Y_{1t} + \phi \sum_{t=1}^T Y_{2t}$ is sufficient for A , and it is symmetric in t .

Proof. The joint density is

$$\prod_{t=1}^T \phi^{z_{2t}} \frac{A^{z_{1t}+z_{2t}}}{\Gamma(z_{1t})\Gamma(z_{2t})} y_{1t}^{z_{1t}-1} y_{2t}^{z_{2t}-1} e^{-A(y_{1t}+\phi y_{2t})} = e^{-A s_\theta(y)} A^{\sum_t (z_{1t}+z_{2t})} h(y),$$

with $s_\theta(y) = \sum_t y_{1t} + \phi \sum_t y_{2t}$ and $h(y) = \prod_t \phi^{z_{2t}} \Gamma(z_{1t})^{-1} \Gamma(z_{2t})^{-1} y_{1t}^{z_{1t}-1} y_{2t}^{z_{2t}-1}$ free of A . The first two factors depend on y only through $s_\theta(y)$, so by the factorization theorem $s_\theta(Y)$ is sufficient for A ; being a sum over t , it is symmetric. Unlike the canonical statistic in (28), s_θ depends on θ through ϕ . ■

By Lemma 3, $S_{y_0}^\sigma = p^\sigma/p$ is free of A . In the ratio, the factors $\phi^{z_{2t}}$, $A^{z_{1t}+z_{2t}}$ and the gamma functions cancel period by period; the A -coupling cancels because $\sum_t y_{1,\sigma(t-1)} = \sum_t y_{1t}$ and $\sum_t y_{2,\sigma(t-1)} = \sum_t y_{2t}$, so that ϕ disappears together with A ; and the exponents -1 on y_{1t} and y_{2t} cancel by the same symmetry. What remains is (32), which depends on θ only through (β_1, β_2) .

Proposition 4 (MOMENT CONDITIONS FOR THE BIVARIATE GAMMA MODEL). *For the dynamic model (31) with $T = 3$, the moment condition (33) holds for every bounded g , with $S_{Y_0}^{\sigma_1}$ given by (32).*

Proof. The argument is identical to the proof of Proposition 3, with the bivariate gamma transition density $f(y_t, z_t, A)$ (the product of the two gamma densities) in place of the period density. Under the reverse-cyclic permutation, $p^{\sigma_1} = f(y_2, z_1, A) f(y_3, z_2, A) f(y_1, z_3, A)$, with $z_1 = z(Y_0)$ constant given Y_0 , $z_2 = z(y_1)$ depending on y_1 but not y_3 , and $z_3 = z(y_2)$ depending on y_2 but not y_1 . Integrating out $y_3 = (y_{13}, y_{23})$ and then $y_1 = (y_{11}, y_{21})$ each contributes unity, and the remaining integral equals $\mathbb{E}[g(Y_1) \mid Y_0, X, A]$, since under the true model $Y_1 \mid Y_0, X, A$ has density $f(\cdot, z_1, A)$. ■

C.1 Moments of the linear-normal moment function

The discussion of the linear panel data model in Section 4 noted that the moment function (35) is the exponential of a quadratic form in the Gaussian outcomes. It is therefore heavy-tailed, with conditional moments that exist only when the state-dependence coefficient lies in a bounded interval. This appendix section makes the claim in Footnote 11 precise. The results build on Chapter 3 of Mathai and Provost (1992) which establishes that if $W \sim N(\mu, \Sigma)$ in $\mathbb{R}^n$, C is symmetric, and $b \in \mathbb{R}^n$ then

$$E \left[\exp(W'CW + b'W)^k \right] = \det(I - 2k\Sigma C)^{-1/2} \exp\left(\frac{1}{2} v_k' P_k^{-1} v_k - \frac{1}{2} \mu' \Sigma^{-1} \mu\right), \quad (56)$$

where

$$P_k = \Sigma^{-1} - 2kC, \quad v_k = kb + \Sigma^{-1}\mu.$$

The k -th moment is finite if and only if $\Sigma^{-1} - 2kC$ is positive definite.

Take the linear index $z_t(y_{t-1}, x, \theta) = \gamma y_{t-1} + x_t' \beta$, so that (34) is the Gaussian first-order autoregressive model with a fixed effect, and let $T = 3$ with σ_1 the reverse cycle. Writing $S_{y_0}^{\sigma_1} = \exp(Q(y)/\sigma^2)$, expression (35) gives

$$Q(y) = z_1(y_2 - y_1) + z_2(y_3 - y_2) + z_3(y_1 - y_3), \quad \text{with} \quad z_t = \gamma y_{t-1} + x_t' \beta. \quad (57)$$

Expanding and collecting terms, the cross terms in $y_1 y_2$ cancel, leaving

$$Q(y) = \underbrace{\gamma y_3(y_1 - y_2)}_{\text{quadratic}} + \underbrace{\gamma y_0(y_2 - y_1) + \beta' [x_1(y_2 - y_1) + x_2(y_3 - y_2) + x_3(y_1 - y_3)]}_{\text{linear in } y}. \quad (58)$$

Conditional on (Y_0, X, A) , the recursion $Y_t = \gamma Y_{t-1} + x_t' \beta + A + \varepsilon_t$ reads $JY = \mu_0 + \varepsilon$ with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_3)$, so Y is normal with finite mean and variance $\sigma^2 P^{-1}$, where

$$P = \begin{pmatrix} 1 + \gamma^2 & -\gamma & 0 \\ -\gamma & 1 + \gamma^2 & -\gamma \\ 0 & -\gamma & 1 \end{pmatrix}.$$

Writing M for the Hessian of Q , so that its quadratic part is $\frac{1}{2} y' M y$,

$$M = \begin{pmatrix} 0 & 0 & \gamma \\ 0 & 0 & -\gamma \\ \gamma & -\gamma & 0 \end{pmatrix},$$

and noting that Q has no term that is constant in y , we can write $(S_{Y_0}^{\sigma_1})^k = \exp(Y' \frac{1}{2\sigma^2} M Y + b' Y)^k$, with $b \in \mathbb{R}^3$ collecting the linear coefficients of Q/σ^2 . This is exactly the form to which the result in (56) applies, with $\Sigma = \sigma^2 P^{-1}$ and $C = \frac{1}{2\sigma^2} M$. Equation (56) then tells us that the moment is finite when $\Sigma^{-1} - 2k \cdot \frac{1}{2\sigma^2} M = \sigma^{-2} (P - kM)$ is positive definite, i.e. when $P - kM$ is positive definite. Finiteness is therefore governed by γ alone, and not by σ^2 , A , Y_0 , or X .

It remains to turn positive definiteness of $P - kM$ into a condition on γ . The two leading principal minors of $P - kM$ are $1 + \gamma^2$ and $1 + \gamma^2 + \gamma^4$. So positive definiteness comes down to the sign of the full determinant. A direct computation gives

$$\det(P - kM) = 1 - 2k(k - 1) \gamma^2 (1 - \gamma + \gamma^2),$$

so $\mathbb{E}[(S_{Y_0}^{\sigma_1})^k \mid Y_0, X, A] < \infty$ if and only if $2k(k - 1) \gamma^2 (1 - \gamma + \gamma^2) < 1$.

Because $1 - \gamma + \gamma^2 = (\gamma - \frac{1}{2})^2 + \frac{3}{4} > 0$, the region for γ for which the moment is finite is an interval of γ around the origin that shrinks as k grows. For $k = 1$ the condition is satisfied for all γ ; this is consistent with $\mathbb{E}[S_{Y_0}^{\sigma_1} \mid Y_0, X, A] = 1$. The second moment ($k = 2$) requires γ to lie between the two real roots of $1 - 4\gamma^2 + 4\gamma^3 - 4\gamma^4$, that is $\gamma \in (-0.400, 0.575)$, and the higher moments require the successively smaller intervals

$$k = 3 : \quad \gamma \in (-0.251, 0.326), \quad k = 4 : \quad \gamma \in (0.184, 0.224).$$

D Proof of Proposition 1

Theorem 3, stated in Subsection D.1, restates Theorem 4 of Gale and Nikaido (1965) and is included for convenience. This result is the key tool used to establish point identification of θ_0 for our empirical specification (4) in Proposition 1. Lemma 6 complements the theorem by providing simple sufficient conditions for a matrix to be a P -matrix, the main condition required to apply Theorem 3. Subsection D.2 then gives a detailed proof of Proposition 1.

D.1 Auxiliary results

Theorem 3 (THEOREM 4 IN GALE AND NIKAIKO (1965)). *If $\Xi : \Theta \mapsto \mathbb{R}^{Q+1}$, where Θ is a closed rectangular region of $\mathbb{R}^{Q+1}$, is a differentiable mapping such that the jacobian matrix $\nabla \Xi(\theta)$ is a P -matrix for all $\theta \in \Theta$, then Ξ is univalent in Θ .*

Lemma 6 *Let $J \in \mathbb{R}^{n \times n}$. Suppose that (i) $J_{ij} > 0$ for all $i, j \in \{1, \dots, n\}$, and (ii) that J is strictly diagonally dominant, i.e.*

$$\forall i \in \{1, \dots, n\} : \quad J_{ii} > \sum_{j \neq i} J_{ij}$$

Then J is a P -matrix (that is, has all principal minors positive).

Proof. If $J \in \mathbb{R}^{n \times n}$ satisfies (i) and (ii), then it is clear that any principal submatrix of J also satisfies (i) and (ii). Therefore, it suffices to show that (i) and (ii) imply that $\det(J) > 0$.

To this end, it is useful to write $J = D + R$ where $D = \text{diag}(J_{11}, \dots, J_{nn})$. Since $\det(D) = \prod_{i=1}^n J_{ii} > 0$ by (i), D is invertible and we can further write $J = D(I + D^{-1}R)$. Since $\det(J) = \det(D) \times \det(I + D^{-1}R)$, the sign of $\det(J)$ is determined by that of $\det(I + D^{-1}R)$. By (ii), $D^{-1}R$ has row-sums strictly less than 1. Hence, by the Gershgorin disc theorem (e.g Theorem 6.1.1 Horn and Johnson (2012)), all eigenvalues of $D^{-1}R$ have moduli strictly less than one. It follows that any real eigenvalue of $(I + D^{-1}R)$ is strictly positive and thus contributes positively to $\det(I + D^{-1}R)$. Furthermore, since $(I + D^{-1}R) \in \mathbb{R}^{n \times n}$, any of its complex

eigenvalues admits a conjugate pair, and that pair contributes positively to $\det(I + D^{-1}R)$. This establishes that $\det(I + D^{-1}R) > 0$ which concludes the proof. ■

D.2 Proof details

Preliminary calculations. It is helpful to express (37) as

$$\begin{aligned}
& m_{y_0}^{\sigma_1, q}(y, x, \theta) \\
&= \mathbb{1}\{y_1 = q, y_2 = q, y_3 = q\} S_{y_0}^{\sigma_1}(y, x, \theta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} S_{y_0}^{\sigma_1}(y, x, \theta) \\
&+ \sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = l\} S_{y_0}^{\sigma_1}(y, x, \theta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = q\} S_{y_0}^{\sigma_1}(y, x, \theta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = q, y_2 = q, y_3 = r\} S_{y_0}^{\sigma_1}(y, x, \theta) \\
&- \mathbb{1}\{y_1 = q\}
\end{aligned}$$

and to note that, for $y_2 = q$,

$$\begin{aligned}
& S_{y_0}^{\sigma_1}(y, x, \theta) \\
&= \exp(z_3(y_1, q, x, \theta) - z_1(y_1, y_0, x, \theta) + z_1(q, y_0, x, \theta) - z_2(q, y_1, x, \theta) + z_2(y_3, y_1, x, \theta) - z_3(y_3, q, x, \theta)) \\
&= \exp((x_{31, y_1} + x_{12, q} + x_{23, y_3}) \beta + \gamma_q (\mathbb{1}\{y_0 = q\} - \mathbb{1}\{y_3 = q\}) - \gamma_{y_0} \mathbb{1}\{y_1 = y_0\} + \gamma_{y_1} \mathbb{1}\{y_3 = y_1\})
\end{aligned}$$

Thus, for the initial condition $y_0 = q$, we get

$$\begin{aligned}
m_q^{\sigma_1, q}(y, x, \theta) &= \mathbb{1}\{y_1 = q, y_2 = q, y_3 = q\} - \mathbb{1}\{y_1 = q\} \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} \exp((x_{12,q} + x_{21,r})\beta + \gamma_q + \gamma_r) \\
&+ \sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = l\} \exp((x_{31,r} + x_{12,q} + x_{23,l})\beta + \gamma_q) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = q\} \exp((x_{13,q} + x_{31,r})\beta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = q, y_2 = q, y_3 = r\} \exp((x_{32,q} + x_{23,r})\beta)
\end{aligned} \tag{59}$$

As a result,

$$\begin{aligned}
&\frac{\partial m_q^{\sigma_1, q}(y, x, \theta)}{\partial \beta} \\
&= \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} (x_{12,q} + x_{21,r}) \exp((x_{12,q} + x_{21,r})\beta + \gamma_q + \gamma_r) \\
&+ \sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = l\} (x_{31,r} + x_{12,q} + x_{23,l}) \exp((x_{31,r} + x_{12,q} + x_{23,l})\beta + \gamma_q) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = q\} (x_{13,q} + x_{31,r}) \exp((x_{13,q} + x_{31,r})\beta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = q, y_2 = q, y_3 = r\} (x_{32,q} + x_{23,r}) \exp((x_{32,q} + x_{23,r})\beta)
\end{aligned} \tag{60}$$

$$\begin{aligned}
\frac{\partial m_q^{\sigma_1, q}(y, x, \theta)}{\partial \gamma_q} &= \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} \exp((x_{12, q} + x_{21, r})\beta + \gamma_q + \gamma_r) \\
&+ \sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = l\} \exp((x_{31, r} + x_{12, q} + x_{23, l})\beta + \gamma_q)
\end{aligned} \tag{61}$$

and for $r \neq q$

$$\frac{\partial m_q^{\sigma_1, q}(y, x, \theta)}{\partial \gamma_r} = \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} \exp((x_{12, q} + x_{21, r})\beta + \gamma_q + \gamma_r) \tag{62}$$

Constructing a P-matrix. Let $\Xi^{1, \beta}(\theta) = \mathbb{E} \left[m_{y_0}^{\sigma_1, 1}(Y, X, \theta) | Y_0 = 1, X \in \mathcal{X}_1^\beta \right]$ and for $q = 1 \dots, Q$, let $\Xi_{\epsilon_q}^{q, \gamma}(\theta) = \mathbb{E} \left[m_{y_0}^{\sigma_1, q}(Y, X, \theta) | Y_0 = q, X \in \mathcal{X}_{q, \epsilon_q}^\gamma \right]$. Let

$$\Xi(\theta) = \begin{pmatrix} \Xi^{1, \beta}(\theta) \\ \Xi_{\epsilon_1}^{1, \gamma}(\theta) \\ \vdots \\ \Xi_{\epsilon_Q}^{Q, \gamma}(\theta) \end{pmatrix}$$

We want to show that $\forall \theta \in \Theta$, $\nabla \Xi(\theta)$ is a P -matrix for an appropriate selection of positive constants $\epsilon_1, \dots, \epsilon_Q$. The first observation is that Assumptions 1 and 2, conditions (38) and (39), together with the definitions of the sets $\mathcal{X}_{q, \epsilon}^\gamma$ and $\mathcal{X}_q^\beta$ ensure that all entries of $\nabla \Xi(\theta)$ are strictly positive. This is easily seen by inspecting equations (60), (61), (62). Hence by Lemma 6, all we need to show is that $\nabla \Xi(\theta)$ is strictly row diagonally dominant.

From (60),(61),(62), we have

$$\begin{aligned}
& \frac{\partial m_q^{\sigma_{1,q}}(y, x, \theta)}{\partial \beta} - \frac{\partial m_q^{\sigma_{1,q}}(y, x, \theta)}{\partial \gamma_q} - \sum_{r \neq q} \frac{\partial m_q^{\sigma_{1,q}}(y, x, \theta)}{\partial \gamma_r} \\
&= + \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = r\} (x_{12,q} + x_{21,r} - 2) \exp((x_{12,q} + x_{21,r})\beta + \gamma_q + \gamma_r) \\
&+ \sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = l\} (x_{31,r} + x_{12,q} + x_{23,l} - 1) \exp((x_{31,r} + x_{12,q} + x_{23,l})\beta + \gamma_q) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = r, y_2 = q, y_3 = q\} (x_{13,q} + x_{31,r}) \exp((x_{13,q} + x_{31,r})\beta) \\
&+ \sum_{\substack{r=1 \\ r \neq q}}^Q \mathbb{1}\{y_1 = q, y_2 = q, y_3 = r\} (x_{32,q} + x_{23,r}) \exp((x_{32,q} + x_{23,r})\beta)
\end{aligned}$$

For any $x \in \mathcal{X}_q^\beta$, $x_{12,q} + x_{21,r} - 2 > 0$ and $(x_{31,r} + x_{12,q} + x_{23,l} - 1) > 0$ in addition to $(x_{13,q} + x_{31,r}) > 0$ and $(x_{32,q} + x_{23,r}) > 0$. It readily follows, choosing $q = 1$, that

$$\nabla_\beta \Xi^{1,\beta}(\theta) > \sum_{l=1}^Q \nabla_{\gamma_l} \Xi^{1,\beta}(\theta) \quad (63)$$

which means that the first row of $\nabla \Xi(\theta)$ satisfies the condition of strict diagonal dominance.

Next, we would like to show that for all $q \in \{1, \dots, Q\}$:

$$\nabla_{\gamma_q} \Xi_{\epsilon_q}^{q,\gamma}(\theta) > \nabla_\beta \Xi_{\epsilon_q}^{q,\gamma}(\theta) + \sum_{r \neq q} \nabla_{\gamma_r} \Xi_{\epsilon_q}^{q,\gamma}(\theta) \quad (64)$$

for appropriate constants $\epsilon_q > 0$. Let

$$\begin{aligned} \Delta_{q,\epsilon_q}(\theta) &= \nabla_{\gamma_q} \Xi_{\epsilon_q}^{q,\gamma}(\theta) - \sum_{r \neq q} \nabla_{\gamma_r} \Xi_{\epsilon_q}^{q,\gamma}(\theta) \\ &= \mathbb{E} \left[\sum_{\substack{r=1, l=1 \\ r \neq q, l \neq q, r \neq l}}^Q P(Y_1 = r, Y_2 = q, Y_3 = l | Y_0 = q, X) \exp((X_{31,r} + X_{12,q} + X_{23,l})\beta + \gamma_q) | Y_0 = q, X \in \mathcal{X}_{q,\epsilon_q}^\gamma \right] \end{aligned}$$

and

$$\begin{aligned} \nu_{q,\epsilon_q}(\theta) &= \mathbb{E} \left[\sum_{\substack{r=1 \\ r \neq q}}^Q P(Y_1 = r, Y_2 = q, Y_3 = r | Y_0 = q, X) \exp((X_{12,q} + X_{21,r})\beta + \gamma_q + \gamma_r) \right. \\ &\quad + \sum_{\substack{r=1 \\ r \neq q}}^Q P(Y_1 = r, Y_2 = q, Y_3 = q | Y_0 = q, X) \exp((X_{13,q} + X_{31,r})\beta) \\ &\quad \left. + \sum_{\substack{r=1 \\ r \neq q}}^Q P(Y_1 = q, Y_2 = q, Y_3 = r | Y_0 = q, X) \exp((X_{32,q} + X_{23,r})\beta) | Y_0 = q, X \in \mathcal{X}_{q,\epsilon_q}^\gamma \right] \end{aligned}$$

By (60) and the definition of $\mathcal{X}_{q,\epsilon_q}^\gamma$, we have $\nabla_{\beta} \Xi_{\epsilon_q}^{q,\gamma}(\theta) < 2\epsilon_q \nu_{q,\epsilon_q}(\theta) + 3\epsilon_q \Delta_{q,\epsilon_q}(\theta)$. Strict row diagonal dominance will then follow if we can justify the existence of ϵ_q such that $\nabla_{\beta} \Xi_{\epsilon_q}^{q,\gamma}(\theta) < 2\epsilon_q \nu_{q,\epsilon_q}(\theta) + 3\epsilon_q \Delta_{q,\epsilon_q}(\theta) < \Delta_{q,\epsilon_q}(\theta)$, i.e that

$$2\epsilon_q \nu_{q,\epsilon_q}(\theta) < (1 - 3\epsilon_q) \Delta_{q,\epsilon_q}(\theta) \quad (65)$$

To this end, we note the upper bound

$$2\epsilon_q \nu_{q,\epsilon_q}(\theta) \leq 2\epsilon_q (Q - 1) e^{2\epsilon_q \max\{|\underline{\beta}|, |\bar{\beta}|\}} (e^{2\bar{\gamma}} + 2) \quad (66)$$

and the lower bound

$$\Delta_{q,\epsilon_q}(\theta) \geq P(Y_1 \neq q, Y_2 = q, Y_3 \neq q, Y_1 \neq Y_3 \mid Y_0 = q, X \in \mathcal{X}_{q,\epsilon_q}^\gamma) \exp(3\epsilon_q \underline{\beta} + \underline{\gamma})$$

We now proceed to show that $P(Y_1 \neq q, Y_2 = q, Y_3 \neq q, Y_1 \neq Y_3 \mid Y_0 = q, X \in \mathcal{X}_{q, \epsilon_q}^\gamma)$ is uniformly bounded away from zero. First, remark that for $x \in \mathcal{X}_{q, \epsilon_q}^\gamma$,

$$\begin{aligned} & P(Y_1 \neq q, Y_2 = q, Y_3 \neq q, Y_1 \neq Y_3 \mid Y_0 = q, X = x, A = \alpha) \\ & \geq P(Y_1 = r, Y_2 = q, Y_3 = l \mid Y_0 = q, X = x, A = \alpha) \end{aligned}$$

for any $r \neq l, r \neq q, l \neq q$. Fix such a pair (r, l) . This is possible because we assumed $Q \geq 3$. Now, since $x \in \mathcal{X}_{q, \epsilon_q}^\gamma$, $\|x\|_\infty < B$. Furthermore, Θ is compact by Assumption 1. Together with Assumption 2, the multinomial logit specification (7), and (4), this implies that on the compact region $\mathcal{A}_q^{reg}$, there exists a constant $\kappa_q > 0$ such that

$$P(Y_1 = r, Y_2 = q, Y_3 = l \mid Y_0 = q, X = x, A = \alpha) \geq \kappa_q$$

By Assumption 2, there also exists $\rho_{y_0} > 0$ such that for all $x \in \mathcal{X}_{y_0}^{reg}$, $P(A \in \mathcal{A}_{y_0}^{reg} \mid Y_0 = y_0, X = x) \geq \rho_{y_0}$. Averaging over α , we then get that $\forall x \in \mathcal{X}_{q, \epsilon_q}^\gamma \subset \mathcal{X}_q^{reg}$,

$$P(Y_1 = r, Y_2 = q, Y_3 = l \mid Y_0 = q, X = x) \geq \kappa_q \rho_q$$

where $\kappa_q \rho_q$ does not depend on ϵ_q . Finally, by condition (39), $P(Y_0 = q, X \in \mathcal{X}_{q, \epsilon}^\gamma) > 0$ so

$$\begin{aligned} & P(Y_1 \neq q, Y_2 = q, Y_3 \neq q, Y_1 \neq Y_3 \mid Y_0 = q, X \in \mathcal{X}_{q, \epsilon}^\gamma) \\ & \geq P(Y_1 = r, Y_2 = q, Y_3 = l \mid Y_0 = q, X \in \mathcal{X}_{q, \epsilon}^\gamma) \\ & = \mathbb{E} \left[P(Y_1 = r, Y_2 = q, Y_3 = l \mid Y_0 = q, X = x) \mid Y_0 = q, X \in \mathcal{X}_{q, \epsilon}^\gamma \right] \\ & \geq \kappa_q \rho_q > 0 \end{aligned}$$

which in turn implies

$$\Delta_{q, \epsilon_q}(\theta) \geq \kappa_q \rho_q \exp \left(3\epsilon_q \underline{\beta} + \underline{\gamma} \right) \quad (67)$$

We now observe that $\lim_{\epsilon_q \rightarrow 0} 2\epsilon_q \nu_{q, \epsilon_q}(\theta) = 0$ by (66) while $\lim_{\epsilon_q \rightarrow 0} (1 - 3\epsilon_q) \Delta_{q, \epsilon_q}(\theta) \geq \kappa_q \rho_q \exp(\underline{\gamma}) > 0$

by (67). Hence, it is clear that there exists $\epsilon_q > 0$ sufficiently small such that (65) is satisfied for all $\theta \in \Theta$.

Taking stock, we have shown that we can select constants $\epsilon_1 \dots, \epsilon_Q > 0$ such that (64) holds for every $q = 1 \dots, Q$. Henceforth, we fix these appropriate constants. Together with (63), this shows that $\nabla \Xi(\theta)$ is strictly row diagonally dominant with positive entries for any $\theta \in \Theta$. Therefore, $\nabla \Xi(\theta)$ is a P -matrix by Lemma 6, and Theorem 3 guarantees that $\Xi(\theta)$ is univalent. To conclude, we note that by Theorem 1, $\Xi(\theta_0) = 0$. Consequently, the univalence of Ξ ensures that θ_0 is the unique root of the $(Q + 1)$ -dimensional system of equations $\Xi(\theta) = 0$.