Supplemental Appendix: "Transition Probabilities and Moment Restrictions in Dynamic Fixed Effects Logit Models"

Kevin Dano
Princeton University

November 5, 2025

1 Supplementary lemma

Lemma 8. *Let* $n \in \mathbb{N}^*$ *and let* $(u_1, \ldots, u_n) \in \mathbb{R}^n_{++}$ *with* $u_i \neq u_j$ *for* $i \neq j$*. Then,* $\mathcal{F} = \left\{1, \left(\frac{1}{1+u_i Z}\right)_{i=1}^{K}, \left(\frac{u_i Z}{1+u_i Z}\right)_{i=K+1}^{n}\right\}$ *, with* $n > K$*, is linearly independent in the vector space of real-valued rational functions of* $Z \in \mathbb{R}_+$

Proof. Since $\frac{u_i Z}{1+u_i Z} = 1 - \frac{1}{1+u_i Z}$, it is equivalent to prove the linear independence of $\left\{1, \left(\frac{1}{1+u_i Z}\right)_{i=1}^{n}\right\}$. To this end, let $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{1+n}$ such that

$$\lambda_0 + \sum_{j=1}^{n} \lambda_j \frac{1}{1 + u_j z} = 0, \quad \forall z \in \mathbb{R}_+$$

we want to show $\lambda = 0$. It is helpful to rewrite the above in the form

$$\mu_0 + \sum_{j=1}^{n} \mu_j \frac{1}{z + v_j} = 0, \quad \forall z \in \mathbb{R}_+$$

where $\mu_0 = \lambda_0$, $\mu_j = \lambda_j u_j^{-1}, v_j = u_j^{-1}, j = 1, \ldots, n$. Evaluating this expression at arbitrary $z_1 < z_2 < \ldots < z_{n+1}$, positive reals, yields the $(n+1) \times (n+1)$ system of equations: $M\mu = 0$, where $\mu = (\mu_0, \ldots, \mu_n)$, $M = [\iota_{n+1}, C]$, ι_{n+1} is the $(n+1)$-vector of ones, and $C = \left[(z_i + v_j)^{-1}\right]_{\substack{i=1,\ldots,n+1 \\ j=1,\ldots,n}}$ is a Cauchy matrix. Now, a Laplace expansion of M along the first row yields $\det(M) = \sum_{k=1}^{n+1} (-1)^{k+1} \det(C_{k,1})$, where

$C_{k1} = \left[(z_i + v_j)^{-1}\right]_{\substack{i\in\{1,\dots,n+1\}\setminus\{k\} \\ j=1,\dots,n}}$ is the submatrix obtained from removing the k-th row and first column of C. Since $C_{k,1}$ is a $n \times n$ Cauchy matrix for all k, we have (see e.g Horn and Johnson (2012))

$$\det(C_{k,1}) = \frac{\prod_{\substack{i=1 \\ i\neq k}}^{n+1} \prod_{j=1}^{n} (z_i - z_j)(v_i - v_j)}{\prod_{\substack{i=1 \\ i\neq k}}^{n+1} \prod_{j=1}^{n} (z_i + v_j)} > 0,$$

and hence $\det(M) > 0$. Since M is nonsingular, it follows that $\mu = 0$, which in turn implies $\lambda = 0$. □

2 The remaining steps for the AR(p) model with $p > 1$

As indicated in Subsection 4.2, **Step 1)** (b) is entirely analogous to the AR(1) case since the transition probabilities have the same functional form. Thus, as soon as $T \geq p + 2$, we can construct the transition functions displayed in Lemma 9. The proof is identical to that of Lemma 3 in the main text.

Lemma 9. *In model (ARp) with $T \geq p+2$, for all $t \in \{p+1, \dots, T-1\}$, $s \in \{1, \dots, t-p\}$ and $y_1^p \in \mathcal{Y}^p$, let $\mu_s(\theta) = \sum_{r=1}^{p} \gamma_r Y_{is-r} + X_{is}'\beta$, $\kappa_t^{y_1|y_1^p}(\theta) = \sum_{r=1}^{p} \gamma_r y_r + X_{it+1}'\beta$, $\omega_{t,s}^{y_1|y_1^p}(\theta) = \left[1 - e^{(\kappa_t^{y_1|y_1^p}(\theta) - \mu_s(\theta))}\right]^{1-y_1} \left[1 - e^{-(\kappa_t^{y_1|y_1^p}(\theta) - \mu_s(\theta))}\right]^{y_1}$ and define the moment functions:*

$$\zeta_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i) = \mathbb{1}\{Y_{is} = y_1\} + \omega_{t,s}^{y_1|y_1^p}(\theta)\mathbb{1}\{Y_{is} \neq y_1\}\phi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, X_i)$$

Additionally, if $T \geq p+3$, for all collection of ordered indices s_1^J with $J \geq 2$ satisfying $t - p \geq s_1 > \dots > s_J \geq 1$, define analogously

$$\begin{aligned}\zeta_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \dots, Y_{is_J-p}^{s_J}, X_i) &= \mathbb{1}\{Y_{is_J} = y_1\} \\ &\quad + \omega_{t,s_J}^{y_1|y_1^p}(\theta)\mathbb{1}\{Y_{is_J} \neq y_1\}\zeta_\theta^{y_1|y_1^p}(Y_{it-1}^{t+1}, Y_{is_1-p}^{s_1}, \dots, Y_{is_{J-1}-p}^{s_{J-1}}, X_i)\end{aligned}$$

Then

$$\begin{aligned}&\mathbb{E}\left[\zeta_{\theta_0}^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i)|Y_i^0, Y_{i1}^{s-1}, X_i, A_i\right] = \pi_t^{y_1|y_1^p}(A_i, X_i) \\ &\mathbb{E}\left[\zeta_{\theta_0}^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \dots, Y_{is_J-p}^{s_J}, X_i)|Y_i^0, Y_{i1}^{s_J-1}, X_i, A_i\right] = \pi_t^{y_1|y_1^p}(A_i, X_i)\end{aligned}$$

For **Step 2)**, provided $T \geq p+2$, it is clear that the difference between any two distinct transition functions associated to the same transition probability in $t \in \{p+1, \ldots, T-1\}$ will yield a valid moment function. Proposition 2 presents a family of $2^T - (T+1-p)2^p$ valid moment functions generalizing those obtained for the one lag case. Recall that by Theorem 3, this is precisely the total number of moment restrictions for the AR(p) logit model. We verified numerically for various values of T and p that they are linearly independent.

Proposition 2. *In model (ARp)*
if $T \geq p+2$, for all $t \in \{p+1, \ldots, T-1\}$, $s \in \{1, \ldots, t-p\}$ and $y_1^p \in \mathcal{Y}^p$, let

$$\psi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i) = \phi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, X_i) - \zeta_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i),$$

if $T \geq p+3$, for all $t \in \{p+1, \ldots, T-1\}$ and collection of ordered indices s_1^J with $J \geq 2$ satisfying $t-p \geq s_1 > \ldots > s_J \geq 1$, and for all $y_1^p \in \mathcal{Y}^p$, let

$$\psi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i) = \phi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, X_i) - \zeta_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i)$$

Then,

$$\mathbb{E}\left[\psi_{\theta_0}^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i)|Y_i^0, Y_{i1}^{s-1}, X_i, A_i\right] = 0$$
$$\mathbb{E}\left[\psi_{\theta_0}^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i)|Y_i^0, Y_{i1}^{s_J-1}, X_i, A_i\right] = 0$$

The proof is identical to that of Proposition 1. From a practical standpoint, these results show that functional differencing in panel data logit models, i.e solving equation (1) in the main text, can be broken down into a sequence of equivalent simpler subproblems period by period that collectively pin down all moment conditions. Our procedure can be advantageous in sophisticated models with a few lags where numerical/symbolic approaches using the full likelihood as in Honoré and Weidner (2020) may start to prove difficult.

3 An identification argument for AR(p) logit models with $p > 1$

This section discusses ways to leverage our methodology and moment restrictions to identify the common parameters. For ease of exposition, we concentrate on the AR(2) logit model with $T = 4$ and generalize the insights towards the end.

There are two possible paths to inference. The first one is is to consider the "identified set" Θ^I of θ_0 based on the four conditional moment restrictions implied by the AR(2) model:

$$\Theta^I = \left\{ \theta \in \mathbb{R}^{2+K_x} : \mathbb{E}_{\theta_0}\left[\psi_\theta^{y_1|y_1,y_2}(Y_{i0}^4, Y_{i-1}^1, X_i)|Y_i^0, X_i\right] = 0, \quad \forall (y_1, y_2) \in \{0,1\}^2 \right\}$$

and construct confidence sets for θ_0 following e.g Andrews and Shi (2013). Instead, the sharp identified set may be computed following the approach of Dobronyi et al. (2021) if the covariates X_i are discrete with finite support. Alternatively, a second approach which we develop further here is to formulate sensible restrictions on covariates that secure point identification in the spirit of Honoré and Kyriazidou (2000). Specifically, we consider the case where a continuous scalar component W_{i2} of X_{i2} has unbounded positive support conditional on Y_i^0, the other regressors, A_i and has a non-trivial effect β_{0W} of known sign to the econometrician. This is the content of Assumption 1 in which $Z_i = (R_i', W_{i1}, W_{i3}, W_{i4})$, and $X_{it} = (W_{it}, R_{it}') \in \mathbb{R}^{K_x}$ for all $t \in \{1,2,3,4\}$. Dobronyi et al. (2023) used a similar device to develop an alternative distribution-free semiparametric estimator to that of Honoré and Kyriazidou (2000) that can accommodate time effects in the baseline one lag model.

Assumption 1. *(i) The covariate W_{i2} is continuously distributed with unbounded support on $\mathbb{R}_+$ conditional on Y_i^0, Z_i, A_i and (ii) β_{0W} is known to be strictly negative.*

Besides being a technical convenience, Assumption 1 may be reasonable in some situations, e.g in the context of our empirical application, the econometrician may have a confident prior that drug prices affect individual drug consumption negatively. We point out that nothing in the discussion that follows hinges critically on $\beta_W < 0$ and or W_{i2} having support on the positive reals. A set of perfectly symmetric arguments will deliver the same conclusions if instead $\beta_W > 0$ and W_{i2} has unbounded support on $\mathbb{R}_-$.

Assumption 2. *(i) $\theta_0 = (\gamma_{01}, \gamma_{02}, \beta_0')' \in \mathbb{G}_1 \times \mathbb{G}_2 \times \mathbb{B} = \Theta$, $\mathbb{G}_1, \mathbb{G}_2, \mathbb{B}$ compact. The conditional densities of A_i and Z_i verify:*

(ii) $\lim_{w_2 \to \infty} q(a|y^0, z, w_2) = q_\infty(a|y^0, z)$, $\lim_{w_2 \to \infty} p(z|y^0, w_2) = p_\infty(z|y^0)$, *where q_∞ and p_∞ are densities.*

(iii) There exists positive integrable functions $d_0(a), d_1(z), d_2(z)$ such that $q(a|y^0, z, w_2) \leq d_0(a)$ for all $a \in \mathcal{A}$, $d_1(z) \leq p(z|y^0, w_2) \leq d_2(z)$ for all $z \in \mathbb{R}^{K_x - 1}$

(iv) $w_2 \mapsto p(a|y^0, z, w_2), w_2 \mapsto p(z|y^0, w_2)$ *are continuous in w_2.*

Assumption 2 are standard regularity conditions for an application of the dominated convergence theorem that once paired with Assumption 1 are sufficient to establish that θ_0 is *identified at infinity*. The outline of the argument is as follows. Under these assumptions, by sending W_{i2} to ∞, the valid moment function $\psi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i)$ of Proposition 2 reduces to

$$\begin{aligned}\psi_{\theta,\infty}^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, Z_i) = &-(1-Y_{i1})(1-Y_{i2})Y_{i3} \\ &+ \left[e^{X_{i34}'\beta} - 1\right](1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4} \\ &+ e^{-\gamma_1 Y_{i0} + \gamma_2(1-Y_{i-1}) + X_{i31}'\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4} \\ &+ e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X_{i41}'\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})\end{aligned} \tag{3}$$

which occurs because $\lim_{w_2\to\infty} e^{w_2\beta_W} = 0$ and $Y_{i2} = 0$ with probability one conditional on the regressors and the fixed effects. The key observation is that this "limiting" moment function has a similar functional form to the valid moment functions of the AR(1) model with $T = 3$. In turn, this implies monotonicity properties on certain regions of the covariate space that we can exploit to point identify θ_0 in the spirit of Honoré and Weidner (2020). To this end, let $(\bar{x}, \underline{x}) \in \mathbb{R}^2$, such that $\bar{x} > \underline{x}$ and define the sets

$$\begin{aligned}\mathcal{X}_{k,+} &= \{x \in \mathbb{R}^{4K_x} | \bar{x} \geq x_{k,3} \geq x_{k,4} > x_{k,1} \geq \underline{x} \text{ or } \bar{x} \geq x_{k,3} > x_{k,4} \geq x_{k,1} \geq \underline{x}\} \\ \mathcal{X}_{k,-} &= \{x \in \mathbb{R}^{4K_x} | \underline{x} \leq x_{k,3} \leq x_{k,4} < x_{k,1} \leq \bar{x} \text{ or } \underline{x} \leq x_{k,3} < x_{k,4} \leq x_{k,1} \leq \bar{x}\}\end{aligned}$$

for all $k \in \{1, \ldots, K_x\}$. In words, $\mathcal{X}_{k,+}$ is the region of the covariate space in which values of the k-th regressor in periods $t \in \{1, 3, 4\}$ belong to $[\underline{x}, \bar{x}]$ and verify $x_{k,3} \geq x_{k,4} \geq x_{k,1}$ with at least one strict inequality. Instead, $\mathcal{X}_{k,-}$ is the region of the covariate space where realizations of the k-th regressor obey the reverse ranking. With these notations in hands, we have the following theorem,

Theorem 5. *For $T = 4$, suppose that outcomes $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ are generated from model (ARp) with $p = 2$, initial condition $y^0 \in \mathcal{Y}^2$, common parameters $\theta_0 = (\gamma_0', \beta_0') \in \mathbb{R}^{2+K_x}$ and that Assumptions 1 and 2 hold. Further, for all $s \in \{-,+\}^{K_x}$, let $\mathcal{X}_s = \bigcap_{k=1}^{K_x} \mathcal{X}_{k,s_k}$ and suppose that for all $y^0 \in \mathcal{Y}^2$*

$$\lim_{w_2\to\infty} P\left(Y_i^0 = y^0, \quad X_i \in \mathcal{X}_s \,|\, W_{i2} = w_2\right) > 0$$

Let

$$\Psi^{0|0,0}_{s,y^0}(\theta) = \lim_{w_2\to\infty} \mathbb{E}\left[\psi^{0|0,0}_{\theta,\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, X_i) \,|\, Y^0_i = y^0, X_i \in \mathcal{X}_s, W_{i2} = w_2\right]$$

Then, θ_0 is the unique solution to the system of equations

$$\Psi^{0|0,0}_{s,y^0}(\theta) = 0, \quad \forall s \in \{-,+\}^{K_x}, \quad \forall y^0 \in \mathcal{Y}^2$$

Proof. Specializing Proposition 2 to the AR(2) with $T = 4$ yields the valid moment function:

$$\begin{aligned}
&\psi^{0|0,0}_{\theta}(Y_{i4}, Y_{i3}, Y^2_{i-1}, X_i) = \left(e^{\gamma_2 Y_{i0} - X'_{i42}\beta} - 1\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}\\
&+ \left[e^{\gamma_2 Y_{i0} - X'_{i42}\beta} + \left(1 - e^{\gamma_2 Y_{i0} - X'_{i42}\beta}\right) e^{-X'_{i43}\beta} - 1\right](1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}\\
&+ e^{\gamma_1(1-Y_{i0}) + \gamma_2(Y_{i0} - Y_{i-1}) + X'_{i21}\beta} Y_{i1}(1-Y_{i2})Y_{i3}\\
&+ e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X'_{i41}\beta}\left[e^{\gamma_1 + \gamma_2 Y_{i0} - X'_{i42}\beta} + \left(1 - e^{\gamma_1 + \gamma_2 Y_{i0} - X'_{i42}\beta}\right) e^{\gamma_2 - X'_{i43}\beta}\right] Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}\\
&+ e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X'_{i41}\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})\\
&- (1-Y_{i1})Y_{i2}
\end{aligned}$$

Consider, the "limiting" moment function $\psi^{0|0,0}_{\theta,\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i)$ given in (3). For $s \in \{-,+\}^{K_x}$, consider the moment objective

$$\Psi^{0|0,0}_{s,y^0}(\theta) = \lim_{w_2\to\infty} \mathbb{E}\left[\psi^{0|0,0}_{\theta}(Y_{i4}, Y_{i3}, Y^2_{i-1}, X_i) | Y^0_i = y^0, X_i \in \mathcal{X}_s, W_{i2} = w_2\right]$$

We will show in two successive steps (a) and (b) that

$$\begin{aligned}
\Psi^{0|0,0}_{s,y^0}(\theta) &= \lim_{w_2\to\infty} \mathbb{E}\left[\psi^{0|0,0}_{\theta,\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i) | Y^0_i = y^0, X_i \in \mathcal{X}_s, W_{i2} = w_2\right] \quad \text{(a)}\\
&= \mathbb{E}\left[\psi^{0|0,0}_{\theta,\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i) | Y^0_i = y^0, X_i \in \mathcal{X}_s, W_{i2} = \infty\right] \quad \text{(b)}
\end{aligned}$$

To establish (a), we start by observing that the history sequence $(1-Y_{i1})Y_{i2}$ featuring in $\psi^{0|0,0}_{\theta}$ has expectation zero. To see this, note that by iterated expectations

$$\begin{aligned}
&\lim_{w_2\to\infty} \mathbb{E}\left[(1-Y_{i1})Y_{i2} | Y^0_i = y^0, X_i \in \mathcal{X}_s, W_{i2} = w_2\right]\\
&= \lim_{w_2\to\infty} \int \frac{e^{\gamma_{02} y_0 + x'_2\beta_0 + a}}{1 + e^{\gamma_{02} y_0 + x'_2\beta_0 + a}} \frac{1}{1 + e^{\gamma_{01} y_0 + \gamma_{02} y_{i-1} + x'_1\beta_0 + a}} \underbrace{f(a, z | y_0, \mathcal{X}_s, w_2)}_{\text{joint density}} \, da dz
\end{aligned}$$

Now, $f(a, z|y_0, \mathcal{X}_s, w_2) = q(a|y_0, z, w_2)p(z|y_0, \mathcal{X}_s, w_2) = q(a|y_0, z, w_2)\frac{p(z|y_0, w_2)\mathbb{1}\{X_i \in \mathcal{X}_s\}}{\int_{\mathcal{X}_s} p(z|y_0, w_2)dz}$.

Hence, by part (iii) of Assumption 2, an integrable dominating function of the integrand is

$$\frac{e^{\gamma_{02}y_0+x_2'\beta_0+a}}{1+e^{\gamma_{02}y_0+x_2'\beta_0+A_i}}\frac{1}{1+e^{\gamma_{01}y_0+\gamma_{02}y_{i-1}+x_1'\beta_0+a}}p(a,z|y_0,\mathcal{X}_s,w_2)\leq d_0(a)\frac{d_2(z)}{\int_{\mathcal{X}_s}d_1(z)dz}$$

Moreover, by parts (ii)-(iii) of Assumption 2 and the Dominated Convergence Theorem,

$$\lim_{w_2\to\infty}f(a,z|y_0,\mathcal{X}_s,w_2)=q_\infty(a|y_0,z)\frac{p_\infty(z|y_0)\mathbb{1}\{X_i\in\mathcal{X}_s\}}{\int_{\mathcal{X}_s}q_\infty(z|y_0)dz}\equiv f_\infty(a,z|y_0,\mathcal{X}_s)$$

Hence another application of the Dominated Convergence Theorem gives

$$\begin{aligned}
&\lim_{w_2\to\infty}\mathbb{E}\left[(1-Y_{i1})Y_{i2}|Y_i^0=y^0,X_i\in\mathcal{X}_s,W_{i2}=w_2\right]\\
&=\int\lim_{w_2\to\infty}\frac{e^{\gamma_{02}y_0+x_2'\beta_0+a}}{1+e^{\gamma_{02}y_0+x_2'\beta_0+a}}\frac{1}{1+e^{\gamma_{01}y_0+\gamma_{02}y_{-1}+x_1'\beta_0+a}}f(a,z|y_0,\mathcal{X}_s,w_2)dadz\\
&=\int 0\times f_\infty(a,z|y_0,\mathcal{X}_s)dadz\\
&=0
\end{aligned}$$

where the third line follows from the fact that $\lim_{w_2\to\infty}e^{w_2\beta_W}=0$ by Assumption 1. Applying the same arguments to each remaining summand of $\psi_\theta^{0|0,0}$ and collecting terms delivers (a). To obtain (b), we note that by part (iv) of Assumption 1, $w_2\mapsto\mathbb{E}\left[\psi_{\theta,\infty}^{0|0,0}(Y_{i4},Y_{i3},Y_{i-1}^2,Z_i)|Y_i^0=y^0,X_i\in\mathcal{X}_s,W_{i2}=w_2\right]$ is continuous with a well defined limit at infinity in light of (a). As a result, we can work directly with its continuous extension at infinity. Let us focus on the initial condition $y_0=y_{-1}=0$. It is clear from Equation (3) that $\Psi_{s,0,0}^{0|0,0}(\theta)$ does not depend on γ_1. Furthermore, by parts (i) of Assumption 2 we note that we have the following integrable dominating functions for the derivative:

$$\left|\frac{\partial\psi^{0|0,0}_{\theta,\infty}(Y_{i4},Y_{i3},Y^2_{i-1},Z_i)}{\partial\gamma_2}\right| = e^{\gamma_2+X'_{i31}\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4} \leq \sup_{g_2\in\mathbb{G}_2,b\in\mathbb{B}} e^{g_2+2\max(|\bar{x}|,|\underline{x}|)\|b\|_1}$$

$$\begin{aligned}\left|\frac{\partial\psi^{0|0,0}_{\theta,\infty}(Y_{i4},Y_{i3},Y^2_{i-1},Z_i)}{\partial\beta_k}\right| &= \Big|X_{ik,34}e^{X'_{i34}\beta}(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}\\ &+ X_{ik,31}e^{\gamma_2+X'_{i31}\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}\\ &+ X_{ik,41}e^{\gamma_2+X'_{i31}\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})\Big|\\ &\leq \big|X_{ik,34}\big|e^{X'_{i34}\beta} + \big|X_{ik,31}\big|e^{\gamma_2+X'_{i31}\beta} + \big|X_{ik,41}\big|e^{\gamma_2+X'_{i31}\beta}\\ &\leq 2\max(|\bar{x}|,|\underline{x}|)\sup_{b\in\mathbb{B}} e^{2\max(|\bar{x}|,|\underline{x}|)\|b\|_1}(1+2\sup_{g_2\in\mathbb{G}_2} e^{g_2})\end{aligned}$$

Hence, by Leibniz integral rule, we get

$$\begin{aligned}\frac{\partial\Psi^{0|0,0}_{s,0,0}(\theta)}{\partial\gamma_2} &= \mathbb{E}\left[\frac{\partial\psi^{0|0,0}_{\theta,\infty}(Y_{i4},Y_{i3},Y^2_{i-1},Z_i)}{\partial\gamma_2}|Y^0_i=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\right]\\ &= \mathbb{E}\left[e^{\gamma_2+X'_{i31}\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}|Y^0_i=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\right]\\ &= \mathbb{E}\left[e^{\gamma_2+X'_{i31}\beta}\underbrace{\mathbb{E}\left[Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}|Y^0_i=(0,0),Z_i,W_{i2}=\infty,A_i\right]}_{>0}|Y^0_i=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\right]\\ &> 0\end{aligned}$$

Similarly,

$$\begin{aligned}&\frac{\partial\Psi^{0|0,0}_{s,0,0}(\theta)}{\partial\beta_k}\\ &= \mathbb{E}\left[\frac{\partial\psi^{0|0,0}_{\theta,\infty}(Y_{i4},Y_{i3},Y^2_{i-1},Z_i)}{\partial\beta_k}|Y^0_i=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\right]\\ &= \mathbb{E}\Big[X_{ik,34}e^{X'_{i34}\beta}\times\\ &\underbrace{\mathbb{E}\left[(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}|Y^0_i=(0,0),Z_i,W_{i2}=\infty,A_i\right]}_{>0}|Y^0_i=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\Big]\\ &+ \mathbb{E}\Big[X_{ik,31}e^{\gamma_2+X'_{i31}\beta}\times\end{aligned}$$

$$\underbrace{\mathbb{E}\left[Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}|Y_i^0=(0,0),Z_i,W_{i2}=\infty,A_i\right]}_{>0}|Y_i^0=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\Bigg]$$

$$+\,\mathbb{E}\Big[X_{ik,41}e^{\gamma_2+X_{i31}'\beta}\times$$

$$\underbrace{\mathbb{E}\left[Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})|Y_i^0=(0,0),Z_i,W_{i2}=\infty,A_i\right]}_{>0}|Y_i^0=(0,0),X_i\in\mathcal{X}_s,W_{i2}=\infty\Bigg]$$

The last display shows that $\frac{\partial\Psi_{s,0,0}^{0|0,0}(\theta)}{\partial\beta_k}>0$ if $s_k=+$ and $\frac{\partial\Psi_{s,0,0}^{0|0,0}(\theta)}{\partial\beta_k}<0$ if $s_k=-$. Therefore, appealing to Lemma 2 in Honoré and Weidner (2020), we conclude that the 2^{K_x} system of equations in K_x+1 unkowns given by:

$$\Psi_{s,0,0}^{0|0,0}(\theta)=0,\quad\forall s\in\{-,+\}^{K_x}$$

has at most one solution. It is precisely (γ_{02},β_0), since the validity of $\psi_\theta^{0|0,0}(Y_{i4},Y_{i3},Y_{i-1}^2,X_i)$ for arbitrary X_i directly implies the validity of the limiting moment $\psi_{\theta,\infty}^{0|0,0}(Y_{i4},Y_{i3},Y_{i-1}^2,Z_i)$ at "$W_{i2}=\infty$". Then, notice that for any other initial condition $y^0\in\{(0,1),(1,0),(1,1)\}$, the objective $\Psi_{s,y^0}^{0|0,0}(\theta)$ is strictly monotonic in γ_1. Hence, given (γ_{02},β_0), it point identifies γ_{01}. This concludes the proof of Theorem 5. □

Theorem 5 shows that point identification of θ_0 is achievable in higher-order dynamic logit models in short panels. The main cost for this guarantee is Assumption 1 which presumes knowledge of the data generating process beyond the baseline setup. Additionally, there should be sufficient variation in the regressors X_{it} as $W_{i2}\mapsto\infty$ to ensure that $\lim_{w_2\to\infty}P\left(Y_i^0=y^0,\quad X_i\in\mathcal{X}_s\,|\,W_{i2}=w_2\right)>0$ for all $s\in\{-,+\}^{K_x}$. Our arguments are easily generalizable to AR(p) models with lag order $p\geq 3$. Under natural extensions of Assumptions 1 and 2, the model parameters $\theta_0=(\gamma_{01},\ldots,\gamma_{0p},\beta_0')$ are *identified at infinity* provided $T\geq 2+p$.

Remark 8 (Time effects)**.** Theorem 5 does not readily deals with time effects since such covariates may violate the condition $\lim_{w_2\to\infty}P\left(Y_i^0=y^0,\quad X_i\in\mathcal{X}_s\,|\,W_{i2}=w_2\right)>0$. For example, it is clear that if the k-th covariate is a time trend, the rank ordering on the regressors demanded by the sets $\mathcal{X}_{k,+},\mathcal{X}_{k,-}$ cannot be satisfied. Nevertheless, it is possible to adapt the argument to accommodate these cases. Suppose for concreteness that $X_{ikt}=t$. By further sending W_{i3} to infinity, the limiting moment function of equation (3) reduces to

$$\psi^{0|0,0}_{\theta,\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i) = -(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}$$
$$+ e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + 3\beta_k + X'_{-k,i41}\beta_{-k}} Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})$$

where we momentarily use the shorthands $X_{-k,i41}$ and β_{-k} to denote all covariates and slope coefficients besides the k-th. For $(Y_{i0}, Y_{i-1}) = (0, 0)$, this valid moment function only depends on β and we can view β_k as playing the role of the state dependence coefficients in the proof of Theorem 5. This means that under additional regularity assumptions, analogous arguments to those in the proof will point identify β_0. Varying the initial condition is then sufficient to point identify γ_0 given the monotonicity of the moment function in (γ_1, γ_2).

4 Proofs for the VAR(1) logit model

Proof of Lemma 4. We have

$$\mathbb{E}\left[\phi^{k|k}_\theta(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y^{t-1}_{i1}, X_i, A_i\right] = P(Y_{it} = k|Y_{i0}, Y^{t-1}_{i1}, X_i, A_i)$$
$$\times \sum_{l\in\mathcal{Y}} P(Y_{it+1} = l|Y_{i0}, Y^{t-1}_{i1}, Y_{it} = k, X_i, A_i)\phi^{k|k}_\theta(l, k, Y_{it-1}, X_i)$$
$$= \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}}}$$
$$\times \sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} \frac{e^{l_m(\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i}}} e^{\sum_{m=1}^M (l_m - k_m)\left(\sum_{j=1}^M \gamma_{mj}(Y_{j,it-1} - k_j) - \Delta X'_{m,it+1}\beta_m\right)}$$
$$= \sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} \frac{e^{l_m(\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i}}} \frac{e^{k_m(\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}}}$$
$$= \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}}} \frac{1}{1 + e^{\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i}}}$$
$$\times \sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} e^{l_m(\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i})}$$

Since

$$\sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} e^{l_m(\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i})} = \prod_{m=1}^{M} (1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}})$$

we finally get

$$\begin{aligned}
&\mathbb{E}\left[\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] \\
&= \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}}} \frac{1}{1 + e^{\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i}}} \\
&\times \prod_{m=1}^{M} (1 + e^{\sum_{j=1}^M \gamma_{mj} Y_{j,it-1} + X'_{m,it}\beta_m + A_{m,i}}) \\
&= \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{mj} k_j + X'_{m,it+1}\beta_m + A_{m,i}}} \\
&= \pi_t^{k|k}(A_i, X_i)
\end{aligned}$$

Proof of Lemma 5. By definition, for $T \geq 3$, and for t, s such that $T-1 \geq t > s \geq 1$:

$$\begin{aligned}
&\mathbb{E}\left[\zeta_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] = P(Y_{is} = k|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i) + \\
&\sum_{l \in \mathcal{Y}\setminus\{k\}} \omega_{t,s,l}^{k|k}(\theta) \mathbb{E}\left[\mathbb{1}\{Y_{is} = l\}\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \\
&= \prod_{m=1}^{M} \frac{e^{k_m(\mu_{m,s}(\theta) + A_{m,i})}}{1 + e^{\mu_{m,s}(\theta) + A_{m,i}}} + \sum_{l \in \mathcal{Y}\setminus\{k\}} \omega_{t,s,l}^{k|k}(\theta) \pi_t^{k|k}(A_i, X_i) P(Y_{is} = l|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i) \\
&= \prod_{m=1}^{M} \frac{e^{k_m(\mu_{m,s}(\theta) + A_{m,i})}}{1 + e^{\mu_{m,s}(\theta) + A_{m,i}}} + \sum_{l \in \mathcal{Y}\setminus\{k\}} \left[1 - e^{\sum_{j=1}^M (l_j - k_j)\left[\kappa_{j,t}^{k|k}(\theta) - \mu_{j,s}(\theta)\right]}\right] \prod_{m=1}^{M} \frac{e^{k_m(\kappa_{m,t}^{k|k}(\theta) + A_{m,i})}}{1 + e^{\kappa_{m,t}^{k|k}(\theta) + A_{m,i}}} \frac{e^{l_m(\mu_{m,s}(\theta) + A_{m,i})}}{1 + e^{\mu_{m,s}(\theta) + A_{m,i}}} \\
&= \prod_{m=1}^{M} \frac{e^{k_m(\kappa_{m,t}^{k|k}(\theta) + A_{m,i})}}{1 + e^{\kappa_{m,t}^{k|k}(\theta) + A_{m,i}}} \\
&= \pi_t^{k|k}(A_i, X_i)
\end{aligned}$$

The first line follows from the measurability of the weight $\omega_{t,s,l}^{k|k}(\theta)$ with respect to the conditioning set and the linearity of conditional expectations. The second line uses the definition of $\mu_{j,s}(\theta)$ and follows from the law of iterated expectations and Lemma 4. The third line makes use of the definition of $\kappa_{m,t}^{k|k}(\theta)$ and $\omega_{t,s,l}^{k|k}(\theta)$ and the penultime line uses Appendix Lemma 7.

5 Moment restrictions for the dynamic multinomial logit model

This section presents new results for the first-order dynamic multinomial logit model with fixed effects, MAR(1) for short. The model is described by

$$Y_{it} = \arg\max_{c \in \mathcal{Y}} Y^*_{ict}, \quad t = 1, \dots, T \tag{MAR1}$$
$$Y^*_{ict} = \sum_{l=0}^{C} \gamma_{cl} \mathbb{1}\{Y_{it-1} = l\} + X'_{ict}\beta_c + A_{ic} + \epsilon_{ict}$$

where $\mathcal{Y} = \{0, 1, \dots, C\}$, ϵ_{ict} are serially independent identically distributed type I extreme value errors, independent of strictly exogenous regressors $X_i = (X_{i1}, \dots, X_{iT}) \in \mathcal{X}^T$, lagged outcomes variables, and of fixed effects $A_i = (A_{i0}, \dots, A_{iC}) \in \mathcal{A}$. Here, we assume $\mathcal{X} \subseteq \mathbb{R}^{K_0} \times \dots \times \mathbb{R}^{K_C}$, $\mathcal{A} = \mathbb{R}^{C+1}$. The initial condition is $Y_{i0} \in \mathcal{Y}$ and as usual we leave its distribution conditional on (A_i, X_i) unrestricted. This setup gives rise to the following transition probabilities:

$$\pi_t^{k|l}(A_i, X_i) = P(Y_{it+1} = k | Y_{it} = l, X, A) = \frac{e^{\gamma_{kl} + X'_{kt+1}\beta_k + A_k}}{\sum_{c=0}^{C} e^{\gamma_{cl} + X'_{ct+1}\beta_j + A_c}}, \quad t = 1, \dots, T$$

Following Magnac (2000), we normalize the transition parameters and fixed effect of the reference alternative "0" to zero [1]. That is $\gamma_{j0} = \gamma_{0j} = 0$ for all $j \in \mathcal{Y}$ and $A_{i0} = 0$, leaving $\theta = \big((\gamma_{kl})_{k,l \geq 1}, (\beta_l)_{l \geq 0}\big)$ as the unknown model parameters.

Magnac (2000) studied the "pure" case without covariates and showed that an extension of the conditional likelihood approach proposed by Chamberlain (1985) can be used to identify and estimate the state-dependence parameters. Honoré and Kyriazidou (2000) showed that this argument carries over to the case with exogenous explanatory variables if one matches the regressors across specific time periods. The two-step approach expounded below offers an alternative path for estimation circumventing the need for matching.

Step 1). Similarly to the VAR(1) model, the MAR(1) seems to admit transition functions only for the probabilities of staying in the same state, namely $\pi_t^{k|k}(A_i, X_i)$ for $k \in \mathcal{Y}$. This feature appears to be a common trait of multidimensional fixed effects specifications. To derive the set of these transition functions available from

[1] The state dependence parameters of the reference state cannot be identified so a normalization constraint must be imposed. Setting $A_{i0} = 0$ is also without loss of generality since we can always redefine the fixed effect as $A^*_{ik} = A_{ik} - A_{i0}$.

$T = 2$, we can apply the reasoning of Subsection 3.2.1 in the main text and seek $\phi_\theta^{k|k}(.), k \in \mathcal{Y}$ satisfying:

$$\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = \mathbb{1}\{Y_{it} = k\}\phi_\theta^{k|k}(Y_{it+1}, k, Y_{it-1})$$
$$\mathbb{E}\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) \mid Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{k|k}(A_i, X_i)$$

Upon obtaining their analytical expressions for the simplest case with $C = 2$, it is easy to conjecture and then verify as we do in the proof below that the generic expressions are as displayed in Lemma 10.

Lemma 10. *In model (MAR1) with $T \geq 2$ and $t \in \{1, \ldots, T-1\}$, let for all $k \in \mathcal{Y}$*

$$\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i) = \mathbb{1}\{Y_{it} = k\}e^{\sum_{c\in\mathcal{Y}\setminus\{k\}} \mathbb{1}\{Y_{it+1}=c\}\left(\sum_{j\in\mathcal{Y}}(\gamma_{cj}-\gamma_{kj})\mathbb{1}(Y_{it-1}=j)+\gamma_{kk}-\gamma_{ck}+\Delta X'_{ikt+1}\beta_k-\Delta X'_{ict+1}\beta_c\right)}$$

Then:

$$\mathbb{E}\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{k|k}(A_i, X_i) = \frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum\limits_{c=0}^{C} e^{\gamma_{ck}+X'_{ict+1}\beta_j+A_{ic}}}$$

Proof. We have:

$$\mathbb{E}\left[\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = P(Y_{it} = k|Y_i^0, Y_{i1}^{t-1}, X_i, A_i)\times$$
$$\sum_{l\in\mathcal{Y}} P(Y_{it+1} = l|Y_i^0, Y_{i1}^{t-1}, Y_{it} = k, X_i, A_i)\phi_\theta^{k|k}(l, k, Y_{it-1}, X_i)$$
$$= \frac{e^{\sum_{c=0}^{C}\gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum\limits_{j=0}^{C} e^{\sum_{c=0}^{C}\gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}} \times \sum_{l\in\mathcal{Y}} \frac{e^{\gamma_{lk}+X'_{ilt+1}\beta_l+A_{il}}}{\sum\limits_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}\phi_\theta^{k|k}(l, k, Y_{it-1}, X_i)$$
$$= \frac{e^{\sum_{c=0}^{C}\gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum\limits_{j=0}^{C} e^{\sum_{c=0}^{C}\gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}}\times$$
$$\left(\frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum\limits_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}} + \sum_{l\in\mathcal{Y}\setminus\{k\}} \frac{e^{\gamma_{lk}+X'_{ilt+1}\beta_l+A_{il}}}{\sum\limits_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}e^{\left(\sum_{j=0}^{C}(\gamma_{lj}-\gamma_{kj})\mathbb{1}(Y_{it-1}=j)+\gamma_{kk}-\gamma_{lk}+\Delta X'_{ikt+1}\beta_k-\Delta X'_{ilt+1}\beta_l\right)}\right)$$
$$= \frac{e^{\sum_{c=0}^{C}\gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum\limits_{j=0}^{C} e^{\sum_{c=0}^{C}\gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}} \times \frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum\limits_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}$$

$$+\frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum_{j=0}^{C} e^{\sum_{c=0}^{C}\gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}} \times \sum_{l\in\mathcal{Y}\setminus\{k\}} \frac{1}{\sum_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}} e^{\sum_{j=0}^{C}\gamma_{lj}\mathbb{1}(Y_{it-1}=j)+X'_{ilt}\beta_l+A_{il}}$$

$$= \frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum_{j=0}^{C} e^{\sum_{c=0}^{C}\gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}} \frac{1}{\sum_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}} \sum_{l\in\mathcal{Y}} e^{\sum_{j=0}^{C}\gamma_{lj}\mathbb{1}(Y_{it-1}=j)+X'_{ilt}\beta_l+A_{il}}$$

$$= \frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum_{j=0}^{C} e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}$$

$$= \pi_t^{k|k}(A_i, X_i)$$

□

From there, if $T \geq 4$, we can draw once again on the partial fraction decomposition of Appendix Lemma 6 to get more transition functions associated to the same transition probabilities. Of course, we reiterate that the (multivariate) rational fraction structure of the transition probabilities when the fixed effects are viewed as polynomials is essential to these results. Lemma 11 gives their expressions and concludes **Step 1)**.

Lemma 11. *In model (MAR1) with $T \geq 3$, for all t, s such that $T-1 \geq t > s \geq 1$, let for all $(c,k) \in \mathcal{Y}^2$: $\mu_{c,s}(\theta) = \sum_{j=1}^{C}\gamma_{cj}\mathbb{1}(Y_{is-1}=j) + X'_{ics}\beta_c - X'_{i0s}\beta_0$, $\kappa_{c,t}^{k|k}(\theta) = \gamma_{ck} + X'_{ict+1}\beta_c - X'_{i0t+1}\beta_0$, $\omega_{t,s,c}^{k|k}(\theta) = 1 - e^{(\kappa_{c,t}^{k|k}(\theta)-\mu_{c,s}(\theta))-(\kappa_{k,t}^{k|k}(\theta)-\mu_{k,s}(\theta))}$ and define the moment functions*

$$\zeta_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i) = \mathbb{1}\{Y_{is}=k\} + \sum_{l\in\mathcal{Y}\setminus\{k\}} \omega_{t,s,l}^{k|k}(\theta)\mathbb{1}\{Y_{is}=l\}\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$$

Additionally, if $T \geq 4$, for any t and ordered collection of indices s_1^J, $J \geq 2$, satisfying $T-1 \geq t > s_1 > \ldots > s_J \geq 1$, define analogously,

$$\begin{aligned}\zeta_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i) &= \mathbb{1}\{Y_{is_J}=k\} \\ &+ \sum_{l\in\mathcal{Y}\setminus\{k\}} \omega_{t,s_J,l}^{k|k}(\theta)\mathbb{1}\{Y_{is_J}=l\}\zeta_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_{j-1}-1}^{s_{J-1}}, X_i)\end{aligned}$$

Then,

$$\mathbb{E}\left[\zeta_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] = \pi_t^{k|k}(A_i, X_i)$$
$$\mathbb{E}\left[\zeta_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i)|Y_{i0}, Y_{i1}^{s_J-1}, X_i, A_i\right] = \pi_t^{k|k}(A_i, X_i)$$

Proof. By construction for $T \geq 3$, and t, s such that $T-1 \geq t > s \geq 1$,

$$
\begin{aligned}
&\mathbb{E}\left[\zeta_{\theta_0}^{0|0}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \\
&= P(Y_{is} = 0|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i) \\
&+ \sum_{l \in \mathcal{Y}\setminus\{0\}} \omega_{t,s,l}^{0|0}(\theta)\mathbb{E}\left[\mathbb{1}\{Y_{is} = l\}\mathbb{E}\left[\phi_\theta^{0|0}(Y_{it-1}^{t+1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right]|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \\
&= \frac{1}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}} + \sum_{l=1}^{C} \omega_{t,s,l}^{0|0}(\theta)\mathbb{E}\left[\mathbb{1}\{Y_{is} = l\}|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right]\pi_t^{0|0}(A_i, X_i) \\
&= \frac{1}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}} + \sum_{l=1}^{C}\left(1 - e^{(\kappa_{l,t}^{0|0}(\theta)-\mu_{l,s}(\theta))}\right)\frac{e^{\mu_{l,s}(\theta)+A_{il}}}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}}\frac{1}{1+\sum_{c=1}^{C} e^{\kappa_{c,t}^{0|0}(\theta)+A_{ic}}} \\
&= \frac{1}{1+\sum_{c=1}^{C} e^{\kappa_{c,t}^{0|0}(\theta)+A_{ic}}} \\
&= \pi_t^{0|0}(A_i, X_i)
\end{aligned}
$$

The first line follows from the measurability of the weight $\omega_{t,s,l}^{0|0}(\theta)$ with respect to the conditioning set and the linearity of conditional expectations. The second line uses the definition of $\mu_{c,s}(\theta)$ and follows from the law of iterated expectations and Lemma 10. The third line makes use of the definition of $\kappa_{c,t}^{0|0}(\theta)$, $\omega_{t,s,l}^{0|0}(\theta)$ and the normalization $\gamma_{c0} = \gamma_{0c} = 0$, $A_{0c} = 0$ for all $c \in \mathcal{Y}$. The penultime line uses Appendix Lemma 6.

Analogously, for all $k \in \mathcal{Y} \setminus \{0\}$,

$$
\begin{aligned}
&\mathbb{E}\left[\zeta_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \\
&= P(Y_{is} = k|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i) \\
&+ \sum_{l \in \mathcal{Y}\setminus\{k\}} \omega_{t,s,l}^{k|k}(\theta)\mathbb{E}\left[\mathbb{1}\{Y_{is} = l\}\mathbb{E}\left[\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right]|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \\
&= \frac{e^{\mu_{k,s}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}} + \sum_{l \in \mathcal{Y}\setminus\{k\}} \omega_{t,s,l}^{k|k}(\theta)\mathbb{E}\left[\mathbb{1}\{Y_{is} = l\}|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right]\pi_t^{k|k}(A_i, X_i) \\
&= \frac{e^{\mu_{k,s}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}} \\
&+ \sum_{l \in \mathcal{Y}\setminus\{k\}}\left(1 - e^{(\kappa_{l,t}^{k|k}(\theta)-\mu_{l,s}(\theta))-(\kappa_{k,t}^{k|k}(\theta)-\mu_{k,s}(\theta))}\right)\frac{e^{\mu_{l,s}(\theta)+A_{il}}}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}}\frac{e^{\kappa_{k,t}^{k|k}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C} e^{\kappa_{c,t}^{k|k}(\theta)+A_{ic}}} \\
&= \frac{e^{\mu_{k,s}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}} + \left(1 - e^{-\kappa_{k,t}^{k|k}(\theta)+\mu_{k,s}(\theta)}\right)\frac{1}{1+\sum_{c=1}^{C} e^{\mu_{c,s}(\theta)+A_{ic}}}\frac{e^{\kappa_{k,t}^{k|k}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C} e^{\kappa_{c,t}^{k|k}(\theta)+A_{ic}}}
\end{aligned}
$$

$$
+\sum_{\substack{l=1\\l\neq k}}^{C}\left(1-e^{(\kappa^{k|k}_{l,t}(\theta)-\mu_{l,s}(\theta))-(\kappa^{k|k}_{k,t}(\theta)-\mu_{k,s}(\theta))}\right)\frac{e^{\mu_{l,s}(\theta)+A_{il}}}{1+\sum_{c=1}^{C}e^{\mu_{c,s}(\theta)+A_{ic}}}\frac{e^{\kappa^{k|k}_{k,t}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C}e^{\kappa^{k|k}_{c,t}(\theta)+A_{ic}}}
$$

$$
=\frac{e^{\kappa^{k|k}_{k,t}(\theta)+A_{ik}}}{1+\sum_{c=1}^{C}e^{\kappa^{k|k}_{c,t}(\theta)+A_{ic}}}
$$

$$
=\pi^{k|k}_t(A_i,X_i)
$$

□

For **Step 2)**, we are free to take the difference of any pair of transition functions yielding the same transition probability to form a valid moment function. A collection of linearly independent moment functions can be constructed following Proposition 1, where we substitute the AR(1) transition functions with their MAR(1) counterparts given in Lemmas 10-11.

6 Moment restrictions for a dynamic network formation model

In this section, we show how our methodology can also be fruitfully applied to analyze dynamic network formation models in the spirit of Graham (2013), Graham (2016).

We are now interested in a setting where the undirected social ties of $N \geq 3$ agents are observed over $t = 0, \ldots, T$ periods. Assume that a large random sample of M such network sequences are available to the econometrician, and that link formation from $t \geq 1$ evolves according to:

$$
D_{ijt} = \mathbb{1}\{\gamma_0 D_{ijt-1} + \beta_0 R_{ijt-1} + X'_{ijt}\delta_0 + A_{ij} - \epsilon_{ijt} \geq 0\} \tag{4}
$$

Here, $D_{ijt} \in \{0,1\}$ encodes the presence or absence of a link between agent i and agent j at time t, $R_{ijt} = \sum_{k=1}^{N} D_{ikt}D_{jkt}$ is the number of friends that i and j have in common at t, $X_{ijt} = X_{jit}$ are strictly exogenous covariates (e.g distances, time effects), $A_{ij} = A_{ji}$ is a dyad-specific, time invariant, unobserved heterogeneity term, and ϵ_{ijt} is an iid logistic error. The parameters of interest are $\gamma_0, \beta_0, \delta_0$ which capture respectively state dependence in friendships, transitivity (i.e the benefits of sharing friends in common), and observable "homophily". Let D_t denote the $N \times N$ matrix of social ties at t with ij-th entry D_{ijt}. Additionally, let $\mathcal{D} = (d^1, \ldots, d^n)$ denote the set of all possible undirected networks of N agents - with self-ties ruled out - with cardinality $n = \binom{N}{2}$. The initial network D_0 at $t = 0$ is allowed to correlate freely with the matrix of fixed effects A with ij-th entry A_{ij} and regressors $X = (X_1, \ldots, X_T)$ where X_t is the matrix of covariates at time t with ij-th entry X_{ijt}.

Graham (2013) studied the special case $N = 3$ without covariates and proved the identification of θ_0 using a conditional likelihood argument. To address the general case with regressors, it is helpful to note that rule (4) is in fact similar to the VAR(1) model studied in the main text. This has immediate implications. First, Lemma 4 in the main text tells us that

$$\begin{aligned}\phi_\theta^{d|d}(D_{t+1}, D_t, D_{t-1}, X) &= \mathbb{1}\{D_t = d\} \\ &\times \exp\left(\sum_{i<j}(D_{ijt+1} - d_{ij})\left(\gamma(D_{ijt-1} - d_{ij}) + \beta(R_{ijt-1} - r_{ij}) - \Delta X'_{ijt+1}\beta\right)\right)\end{aligned}$$

is a transition function for $\pi_t^{d|d}(A, X)$, the transition probability of switching from network d to network d. Formally, this means:

$$\mathbb{E}\left[\phi_{\theta_0}^{d|d}(D_{t+1}, D_{it}, D_{t-1}, X)|D_0, D_1^{t-1}, X, A\right] = \pi_t^{d|d}(A, X) = \prod_{i<j}\frac{e^{d_{ij}(\gamma_0 d_{ij} + \beta_0 r_{ij} + X'_{ijt}\delta_0 + A_{ij})}}{1 + e^{\gamma_0 d_{ij} + \beta_0 r_{ij} + X'_{ijt}\delta_0 + A_{ij}}}$$

Then, if $T \geq 3$, Lemma 5 tells us how to construct another set of transition functions associated to $\pi_t^{d|d}(A, X)$. For example, for t, s such that $T - 1 \geq t > s \geq 1$, we can consider

$$\zeta_\theta^{d|d}(D_{t-1}^{t+1}, D_{s-1}^{s}, X) = \mathbb{1}\{D_s = d\} + \sum_{d' \in \mathcal{D}\setminus\{d\}}\omega_{t,s,d'}^{d|d}(\theta)\mathbb{1}\{D_s = d'\}\phi_\theta^{d|d}(D_{t+1}, D_t, D_{t-1}, X)$$

where $\omega_{t,s,d'}^{d|d}(\theta) = 1 - e^{\sum_{i<j}(d'_{ij} - d_{ij})\left[\kappa_{ij,t}^{d|d}(\theta) - \mu_{ij,s}(\theta)\right]}$ with $\kappa_{ij,t}^{d|d} = \gamma d_{ij} + \beta_0 r_{ij} + X'_{ijt+1}\beta$ and $\mu_{ij,s} = \gamma D_{ijs-1} + \beta R_{ijs-1} + X'_{ijs}\beta$. This allows us to form n valid moment functions for the typical case $T = 3$, given by:

$$m_\theta^d(D_1^3, D_0^1, X) = \phi_\theta^{d|d}(D_3, D_2, D_1, X) - \zeta_\theta^{d|d}(D_1^3, D_0^1, X)$$

In turn, this paves the way for an identification analysis of θ_0 and its estimation by GMM leveraging our sample of M iid networks.

7 Additional results for the empirical application

Alongside the EL estimator, we computed the iterated GMM estimator of Hansen et al. (1996). Starting from an initial consistent estimator $\tilde{\theta}_0$ (we used the equallyweighted GMM estimator), it can be described as

$$\tilde{\theta} = \lim_{s\to\infty} \tilde{\theta}_s$$
$$\tilde{\theta}_s = \arg\min_{\theta} \overline{m}_N(\theta)'\overline{\Omega}_N(\tilde{\theta}_{s-1})^{-1}\overline{m}_N(\theta)$$

where $\overline{m}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} m_\theta(Y_i, Y_i^0, X_i)$ and $\overline{\Omega}_N(\theta) = \frac{1}{N}\sum_{i=1}^{N} m_\theta(Y_i, Y_i^0, X_i)m_\theta(Y_i, Y_i^0, X_i)'$. While asymptotically equivalent to the 2-step GMM estimator, Hansen and Lee (2021) recommended its use over it for a few practical reasons. First, $\tilde{\theta}$ eliminates the arbitrariness in the choice of the initial weight matrix of 2-step GMM estimators. Second, because the iteration sequence is a contraction, each iteration is approximately variance reducing in the sense that: $Var(\hat{\theta}_s) \approx c^2 Var(\hat{\theta}_{s-1})$ for some constant $c < 1$. Table 2 shows that the iterated GMM estimates align closely with those obtained using the empirical likelihood approach.

Table 2: Parameter estimates of the trivariate VAR(1) logit based on NLSY data continued

	Empirical Likelihood			Iterated GMM		
	A (I)	M (II)	HD (III)	A (IV)	M (V)	HD (VI)
γ_{m1}	0.48	-0.06	0.38	0.36	-0.09	0.30
	(0.13)	(0.21)	(0.33)	(0.12)	(0.21)	(0.34)
γ_{m2}	0.29	0.83	0.49	0.18	0.78	0.52
	(0.20)	(0.14)	(0.24)	(0.19)	(0.14)	(0.24)
γ_{m3}	-0.29	0.19	0.48	-0.28	0.10	0.48
	(0.31)	(0.22)	(0.23)	(0.30)	(0.22)	(0.23)
age	0.09	-0.09	0.03	0.05	-0.11	0.08
	(0.05)	(0.07)	(0.10)	(0.05)	(0.07)	(0.09)
college	0.25	0.20	0.31	0.28	0.20	0.30
	(0.14)	(0.15)	(0.26)	(0.12)	(0.15)	(0.26)
LR Test	56.45					
"Wald" Test	54.38					

NOTES: *The convergence criterion for the iterated GMM procedure is* $\left\|\hat{\theta}_{s+1} - \hat{\theta}_s\right\| < 10^{-5}$. *Standard errors are reported in parenthesis. Columns titled "A", "M", "HD" report parameter estimates for the alcohol layer, marijuana layer, and hard-drugs layer of the trivariate VAR(1) logit model.*

References

Andrews, D. W. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica*, 81(2):609–666.

Chamberlain, G. (1985). *Heterogeneity, omitted variable bias, and duration dependence*, page 3–38. Econometric Society Monographs. Cambridge University Press.

Dobronyi, C., Gu, J., et al. (2021). Identification of dynamic panel logit models with fixed effects. *arXiv preprint arXiv:2104.04590*.

Dobronyi, C. R., Ouyang, F., and Yang, T. T. (2023). Revisiting panel data discrete choice models with lagged dependent variables. *arXiv preprint arXiv:2301.09379*.

Graham, B. S. (2013). Comment on “social networks and the identification of peer effects” by paul goldsmith-pinkham and guido w. imbens. *Journal of Business and Economic Statistics*, 31(3):266–270.

Graham, B. S. (2016). Homophily and transitivity in dynamic network formation. Technical report, National Bureau of Economic Research.

Hansen, B. E. and Lee, S. (2021). Inference for iterated gmm under misspecification. *Econometrica*, 89(3):1419–1447.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68(4):839–874.

Honoré, B. E. and Weidner, M. (2020). Moment conditions for dynamic panel logit models with fixed effects. *arXiv preprint arXiv:2005.05942*.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

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.