Transition Probabilities and Identifying Moments in Dynamic Fixed Effects Logit Models

Kevin Dano *

This version: July 4, 2023.
Newest version here.

Abstract

This paper deals with estimation of dynamic discrete choice models. Specifically, we introduce an algebraic approach to derive identifying moments in dynamic logit models with strictly exogenous regressors and additive fixed effects. It is based upon two common features in this class of models. First, many (individual-specific) transition probabilities can be expressed as conditional expectations of functions of the data and common parameters given the initial condition, the regressors and the fixed effects. We call such functions *transition functions*. Second, after enough time periods, multiple transition functions map to the same transition probabilities. This motivates a differencing strategy leveraging the multiplicity of transition functions to produce valid moment conditions in panels of adequate length. We detail the construction of identifying moments in scalar models of arbitrary lag order as well as first-order panel vector autoregressions and dynamic multinomial logit models. A simulation study illustrates the small sample performance of GMM estimators based on our methodology.

1 Introduction

In panel data, the availability of multiple observations per sampling unit introduces the possibility to control for unobserved heterogeneity and dynamic feedback. This feature is paramount to discern the nature of causality in a wide range of economic processes. A prominent example that motivates our study is dynamic discrete choice analysis. Here, the main concern is disentangling "true state dependence", i.e the causal effect of past choices on current outcomes, from spurious correlation induced by unit-specific effects (Heckman (1981)). However, achieving this objective in short panels for nonlinear models is notoriously difficult. The combination of dynamic feedback and individual fixed effects typically leads to estimation challenges due in part to the incidental parameter problem (Neyman and Scott (1948)). There are a few notable exceptions

*Department of Economics, University of California Berkeley. E-mail: kdano@berkeley.edu.
I am grateful to Bryan Graham, Stéphane Bonhomme, Jim Powell, Demian Pouzo, Chris Murris, Bocar Ba, Yassine Sbai Sassi, Nick Gebbia, Tahsin Saffat and the audience at the 2023 IAAE Annual Conference for valuable comments and discussions. All errors are my own. The IAAE travel grant for the 2023 IAAE Conference is gratefully acknowledged.

however for which solutions to the incidental parameter problem exist (Arellano and Honoré (2001) provides a review of some known examples). In the context of dynamic discrete choice, dynamic fixed effects logit (DFEL) models, the focus of the present paper, constitute one such exception.

In the "pure" version of the basic model which abstract from covariates other than a first order lag, Cox (1958), Chamberlain (1985) and Magnac (2000) showed that the autoregressive parameter can be consistently estimated by conditional likelihood. This approach relies on the existence of a sufficient statistic linked to the logistic assumption to eliminate the fixed effect. In an important subsequent paper, Honoré and Kyriazidou (2000) extend this idea to a setting with strictly exogenous regressors and show that the conditional likelihood approach remains viable if one can further condition on the regressors being equal in specific periods. This strategy was also found to be effective in dynamic multinomial logit models (Honoré and Kyriazidou (2000)), panel vector autoregressions (Honoré and Kyriazidou (2019)) and dynamic ordered logit models (Muris et al. (2020). At the same time, it has also been noted that the necessity to be able to "match" the covariates imposes two limitations for the conditional likelihood approach: it inherently rules out time effects and implies rates of convergence slower than $\sqrt{N}$ for continuous explanatory variables. Furthermore, calculations from Honoré and Kyriazidou (2000) suggested that it does not easily extend to models with a higher lag order. These undesirable features have encouraged the development of alternative and more flexible methods of estimation.

Recently, Kitazawa et al. (2013, 2016) and Kitazawa (2022) revisited the AR(1) logit model - autoregressive of order one - and proposed a transformation approach that deals with the fixed effects without restricting the nature of the covariates besides the conventional assumption of strict exogeneity. Their methodology effectively leads to moment conditions that can serve as a basis to estimate the model parameters at $\sqrt{N}$-rate by GMM; even with continuous regressors. In independent work, Honoré and Weidner (2020) also derived moment conditions for the AR(1), AR(2) and AR(3) logit models in panels of specific length using the functional differencing technique of Bonhomme (2012). Their approach is partly numerical and requires softwares such as Mathematica to obtain analytical expressions but has a wider scope of applications, e.g dynamic ordered logit specifications (Honoré et al. (2021)). In another recent paper, Dobronyi et al. (2021), the authors analyze the full likelihood of AR(1) and AR(2) logit models with discrete covariates under a new angle that reveals a connection to the *truncated moment problem* in mathematics. Drawing on well established results in that literature, they derive moment equality and moment inequality restrictions - previously unexploited - that fully characterize the sharp identified set.

In this paper, we contribute to this growing body of research by introducing a new algebraic approach to construct moment equality conditions in DFEL models with additive separability between the fixed effects and the explanatory variables (i.e when fixed effects are heterogeneous "intercepts"). This class of models encompasses most specifications studied in prior work but excludes models with heterogeneous coefficients on lagged outcomes and/or regressors as in Chamberlain (1985) and Browning and Carro (2014). We focus our attention on deriving valid moment functions for AR(p) models with arbitrary lag order $p \geq 1$ as well as first-order panel vector autoregressions and dynamic multinomial logit models (Magnac (2000)). At the heart of our methodology lie two key observations. First, the transition probabilities of logit type models can often be expressed as conditional expectations of functions of observables and common parameters given the initial condition, the regressors and the fixed effects. We refer to these moment functions as *transition functions*. They have the convenient feature of not depending on individual fixed effects. Second and crucially, as soon as $T \geq p+2$, where T denotes the number of observations post initial condition, many transition probabilities in periods $t \in \{p+1, \dots, T-1\}$ are associated to at least two distinct transition functions. Consequently, the econometrician can generate valid moment functions from differences between any two transition functions associated to the same transition probability in panels of adequate length. We elaborate on this process in various examples and use the resulting moment functions to document novel identification results and shed a new light on existing ones. In contrast to Honoré and Weidner (2020) and Dobronyi et al. (2021), the proposed methodology does not rely on solving directly a system of equations based on the model full likelihood and hence scales relatively well with model complexity, i.e T, p or the dimensionality of the fixed effects.

Our work also connects to a line of research studying the identification of features of the distribution of fixed effects in discrete choice models. One branch in this literature has focused on developing general optimization tools to compute sharp numerical bounds on average marginal effects. This includes most notably, the linear programming method of Honoré and Tamer (2006) and the quadratic programming method of Chernozhukov et al. (2013). A second branch in this literature has focused instead on harnessing the specificities of logit models to obtain simple analytical bounds. In static logit models, Davezies et al. (2021) exploit mathematical results on the *moment problem* to formulate sharp bounds on the average partial effects of regressors on outcomes. In DFEL models, Aguirregabiria and Carro (2021) are the first to prove the point identification of average marginal effects in the baseline AR(1) logit model when $T \geq 3$. In related work, Dobronyi et al. (2021) make use of their moment equality and moment inequality restrictions to establish sharp bounds on functionals of the fixed effects such as average marginal effects and average posterior means in AR(1) and AR(2) specifications. We complement these results as a byproduct of our methodology: average marginal effects in AR(p) models, with arbitrary $p \geq 1$ and $T \geq 2+p$ are merely expectations of our transition functions.

The remainder of the paper is organized as follows. Section 2 presents the setting, our working assumptions and states our main objective. Section 3 introduces some terminology and gives an outline of our methodology to construct moment conditions. Section 4 implements our approach in AR(1) and AR(p) logit models with $p > 1$ and discusses identification of the model parameters and average marginal effects. Section 5 gives moment conditions in the VAR(1) and the dynamic multinomial logit model with one lag, henceforth MAR(1). Finally, Section 6 documents the finite sample performance of GMM estimators based on our moment conditions in Monte Carlo simulations. Unless stated otherwise, proofs of the main results are gathered in the Appendix.

2 Setting

Let $i = 1, \ldots, N$ denote a population index and $t = 0, \ldots, T$ an index for time. This paper studies threshold-crossing type models describing a discrete outcome Y_{it} through a latent variable involving lagged outcomes (e.g Y_{it-1}), regressors X_{it}, an individual-specific time-invariant unobservable A_i and a schock ϵ_{it}. We focus our attention on models where A_i is additively separable from the other explanatory variables. An initial condition Y_i^0 comes with each model to allow for dynamics. The common parameter θ_0 is the target of interest and governs the influence of lagged outcomes and the regressors on the contemporaneous outcome. Throughout, we assume that the joint distribution of (Y_i^0, X_i, A_i) is unrestricted where $X_i = (X_{i1}, \ldots, X_{iT})$ and thus refer to A_i as a fixed effect in common with the literature. Finally, ϵ_{it} are assumed to be serially independent logistic shocks, independent of (Y_i^0, X_i, A_i) and independent across individuals, except for the MAR(1) model where they are instead extreme value distributed.

The data available to the econometrician consists of the initial condition Y_i^0, the outcome vector $Y_i = (Y_{i1}, \ldots, Y_{iT})$, and the covariates X_i. Interest centers on the identification and estimation of the structural parameter θ_0 in short panels, i.e for fixed T. To this end, the chief objective of this paper is to show how to construct moment functions $\psi_\theta(Y_i, Y_i^0, X_i)$ free of the fixed effect parameter that are valid in the sense that:

$$E\left[\psi_{\theta_0}(Y_i, Y_i^0, X_i) \,|\, Y_i^0, X_i, A_i\right] = 0 \tag{1}$$

When this is possible, the law of iterated expectations implies the conditional moment:

$$E\left[\psi_{\theta_0}(Y_i, Y_i^0, X_i) \,|\, Y_i^0, X_i\right] = 0$$

which can in turn be leveraged to assess the identifiability of θ_0 and form the basis of a GMM estimation strategy. This is the central idea underlying functional differencing (Bonhomme (2012)) and was applied by Honoré and Weidner (2020) to derive valid moment conditions for a class of dynamic logit models with scalar fixed effects. We borrow the same insight but instead of searching for solutions numerically as Honoré and Weidner (2020) do, we propose a complementary and systematic algebraic approach to recover the model's valid moments [1]. Proceeding in this way has the advantage to shed light on the mechanics implied by the logistic assumption which in turn suggest a blueprint to deal with estimation of general DFEL models. For instance, we are able to characterize the expressions of valid moment functions in AR(p) models for arbitrary $p > 1$ which to the best of our knowledge is a new result in the literature. Moreover, our approach easily carries over to multidimensional fixed effect specifications: VAR(1), dynamic network formation models and the MAR(1) in which searching for moments numerically is cumbersome or intractable.

A brief word on notations. Henceforth, we will use the shorthands, $Y_{it_1}^{t_2} = (Y_{it_1}, \ldots, Y_{it_2})$ to denote a collection of random variables over periods t_1 to t_2 with the convention that $Y_{it_1}^{t_2} = \emptyset$ if $t_1 > t_2$. In a

[1] Dobronyi et al. (2021) and Kitazawa (2022) also have an algebraic approach but our methodologies are very different. The first paper uses the full likelihood of the model and focuses on the AR(1) and special instances of the AR(2) model. The second paper has a transformation approach adapted to the AR(1) model.

similar spirit, we may use the notation $y_1^p = (y_1, \dots, y_p)$ to denote any p-dimensional vector of reals with the convention $y_{t_1}^{t_2} = \emptyset$ for $t_1 > t_2$. We let Δ denote the first-differencing operator so that $\Delta Z_{it} = Z_{it} - Z_{it-1}$ for any random variable Z_{it} and make use of the notation $Z_{its} = Z_{it} - Z_{is}$ for $s \neq t$ to accomodate long differences. Last, $\mathbb{1}\{.\}$ denotes the indicator function. For notational convenience, the conventional zero subscript on the common parameters may at times be omitted in the main text and the proofs when doing so causes no confusion.

3 Outline of an algebraic approach to derive valid moment functions

Throughout, we let $\mathcal{Y}$ denote the support of the outcome of interest Y_{it}. Given an initial condition $y^0 \in \mathcal{Y}^p$, $p \geq 1$ referring to the lag order of the model, and regressors $X_i \in \mathbb{R}^{K_x \times T}$, we denote the (one-period ahead) transition probability in period $t \geq 1$ from state $(l_1^t, y^0) \in \mathcal{Y}^t \times \mathcal{Y}^p$ to state $k \in \mathcal{Y}$ as:

$$\pi_t^{k|l_1^t, y^0}(A_i, X_i) = \pi_t^{k|l_1^t, y^0}(A_i, X_i; \theta_0) \equiv P(Y_{it+1} = k \mid Y_i^0 = y^0, Y_{i1}^t = l_1^t, X_i, A_i)$$

The Markovian structure of the models considered in this paper imply that $\pi_t^{k|l_1^t, y^0}(A_i, X_i)$ does not depend on the entire path of past outcomes but rather on the value of the most recent p outcomes. For instance, in an AR(1) model where $p = 1$, we have:

$$\pi_t^{k|l_1^t, y^0}(A_i, X_i) = P(Y_{it+1} = k \mid Y_i^0 = y^0, Y_{i1}^t = l_1^t, X_i, A_i) = P(Y_{it+1} = k \mid Y_{it} = l_t, X_i, A_i)$$

and thus we will suppress the dependence on $(y^0, l_1, \dots, l_{t-1})$ and write $\pi_t^{k|l_t}(A_i, X_i)$. We proceed analogously for models with higher order lags.

We call a *transition function* associated to a transition probability $\pi_t^{k|l_1^t, y^0}(A_i, X_i)$ any moment function $\phi_\theta^{k|l_1^t, y^0}(Y_i, X_i)$ of the data and the common parameters verifying:

$$E\left[\phi_{\theta_0}^{k|l_1^t, y_0}(Y_i, X_i) \,|\, Y_i^0, X_i, A_i\right] = \pi_t^{k|l_1^t, y_0}(A_i, X_i) \tag{2}$$

Our proposed approach to derive valid moment functions in the sense of equation (1) consists of two steps independently of the number of lags and the dimensionality of individual fixed effects. In **Step 1)**, we compute the model's transition functions. Our procedure requires a minimum of $T = p + 1$ periods of observations to accommodate arbitrary regressors and initial condition. In this case we can produce transition functions associated to the transition probabilities in period $t = p$ and Theorem 2 implies that they are unique. However, as soon as $T \geq p + 2$, we show that it becomes possible to construct distinct transition functions associated to the same transition probabilities in periods $t \in \{p + 1, \dots, T - 1\}$. This opens the door for a differencing strategy leading to valid moment functions that we refer to as **Step 2)**.

Step 1) is unsurprisingly easiest with a single lag and we spend much time discussing this foundational case. For AR(p) logit models with $p > 1$, the computation of transition functions is more involved due to the presence of higher order lags. An important and novel contribution of this paper is to provide a general algorithm to fulfill this objective. Another contribution is to show how to construct sequences of transition functions to achieve **Step 2)** when a sufficient number of time periods is available. To do so we build on ideas from Kitazawa (2022). The ensuing sections demonstrate our two-step approach in scalar and multidimensional fixed effect models.

4 Scalar fixed effect models

4.1 Moment conditions for the AR(1) logit model

For exposition, we begin with the baseline AR(1) logit model with fixed effects often formulated as:

$$Y_{it} = \mathbb{1}\{\gamma_0 Y_{it-1} + X_{it}'\beta_0 + A_i - \epsilon_{it} \geq 0\}, \quad t = 1, \dots, T \tag{3}$$

Here, $\mathcal{Y} = \{0, 1\}$, $\theta_0 = (\gamma_0, \beta_0') \in \mathbb{R} \times \mathbb{R}^{K_x}$, the initial condition Y_i^0 consists of the binary-valued random variable Y_{i0} and $A_i \in \mathbb{R}$.

4.1.1 The purely autoregressive case

In the absence of exogenous regressors, model (3) simplifies to:

$$Y_{it} = \mathbb{1}\{\gamma_0 Y_{it-1} + A_i - \epsilon_{it} \geq 0\}, \quad t = 1, \dots, T \tag{4}$$

which was first introduced by Cox (1958) and then revisited in Chamberlain (1985), Magnac (2000). These papers established the identification of γ_0 for $T \geq 3$ via conditional likelihood based on the insight that $(Y_{i0}, \sum_{t=1}^{T-1} Y_{it}, Y_{iT})$ are sufficient statistics for the fixed effect. As discussed in Section 2, our methodology is conceptually different as we seek to directly construct moment functions verifying equation (1).

For this model, the individual-specific transition probability from state l to state k is time-invariant and given by:

$$\pi^{k|l}(A_i) = P(Y_{it+1} = k | Y_{it} = l, A_i) = \frac{e^{k(\gamma_0 l + A_i)}}{1 + e^{\gamma_0 l + A_i}}, \quad \forall (l, k) \in \mathcal{Y}$$

In **Step 1)**, we focus on deriving the transition functions associated to $\pi^{0|0}(A_i)$ and $\pi^{1|1}(A_i)$. The other two transition probabilities $\pi^{1|0}(A_i)$ and $\pi^{0|1}(A_i)$ are effectively redundant since probabilities sum to one. To do so, our proposal is to look for moment functions of the form $\phi_\theta(Y_{it+1}, Y_{it}, Y_{it-1})$ that: 1) are non-zero only for a single realization of the period t outcome and 2) have a conditional expectation given past outcomes and the fixed effect equal to one of the two target transition probabilities. Formally, we search for $\phi_\theta^{0|0}(.)$ and

$\phi_\theta^{1|1}(.)$, functions of three consecutive outcomes, satisfying

$$\begin{aligned}
\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) &= (1-Y_{it})\phi_\theta^{0|0}(Y_{it+1}, 0, Y_{it-1}) \\
\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}) &= Y_{it}\phi_\theta^{1|1}(Y_{it+1}, 1, Y_{it-1}) \\
E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) \mid Y_{i0}, Y_{i1}^{t-1}, A_i\right] &= \pi^{0|0}(A_i) \\
E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}) \mid Y_{i0}, Y_{i1}^{t-1}, A_i\right] &= \pi^{1|1}(A_i)
\end{aligned}$$

Their expressions are simple and provided in Lemma 1 below.

Lemma 1. *In model (4) with $T \geq 2$ and $t \in \{1, \ldots, T-1\}$, let*

$$\begin{aligned}
\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) &= (1-Y_{it})e^{\gamma Y_{it+1}Y_{it-1}} \\
\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}) &= Y_{it}e^{\gamma(1-Y_{it+1})(1-Y_{it-1})}
\end{aligned}$$

Then:

$$\begin{aligned}
E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) | Y_{i0}, Y_{i1}^{t-1}, A_i\right] &= \pi^{0|0}(A_i) = \frac{1}{1+e^{A_i}} \\
E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}) | Y_{i0}, Y_{i1}^{t-1}, A_i\right] &= \pi^{1|1}(A_i) = \frac{e^{\gamma_0 + A_i}}{1+e^{\gamma_0 + A_i}}
\end{aligned}$$

The proof is omitted as it is a special instance of Lemma 2 which covers the case with strictly exogenous explanatory variables.

One heuristic motivation for searching transition functions of this form is the following. In model (4), a key idea to obtain moment restrictions is to look for an intermediate function $\phi_\theta(Y_{it+1}, Y_{it}, Y_{it-1})$ whose conditional expectation given past outcomes and fixed effect does not depend on the past when evaluated at the true parameter [2]. Indeed, if this is somehow achievable, the difference of $\phi_\theta(Y_{it+1}, Y_{it}, Y_{it-1})$ and $\phi_\theta(Y_{is+1}, Y_{is}, Y_{is-1})$ for $t > s$ provides a natural candidate for a valid moment function. To operationalize this idea, we first note that we can generically write,

$$\begin{aligned}
&E\left[\phi_\theta(Y_{it+1}, Y_{it}, Y_{it-1}) \mid Y_{i0} = y_0, Y_{i1}^{t-1} = y_1^{t-1}, A_i = a\right] = \\
&\sum_{y_{t+1}=0}^{1}\sum_{y_t=0}^{1} P(Y_{it+1} = y_{t+1}|Y_{it} = y_t, A_i = a)P(Y_{it} = y_t|Y_{it-1} = y_{t-1}, A_i = a)\phi_\theta(y_{t+1}, y_t, y_{t-1}) \\
&= \frac{e^{\gamma_0 y_{t-1}+a}}{1+e^{\gamma_0 y_{t-1}+a}}\left(\frac{e^{\gamma_0+a}}{1+e^{\gamma_0+a}}\phi_\theta(1,1,y_{t-1}) + \frac{1}{1+e^{\gamma_0+a}}\phi_\theta(0,1,y_{t-1})\right) \\
&+ \frac{1}{1+e^{\gamma_0 y_{t-1}+a}}\left(\frac{e^{a}}{1+e^{a}}\phi_\theta(1,0,y_{t-1}) + \frac{1}{1+e^{a}}\phi_\theta(0,0,y_{t-1})\right)
\end{aligned}$$

Then, the hope is that we can select, $\phi_\theta(1,1,y_{t-1}), \phi_\theta(0,1,y_{t-1}), \phi_\theta(1,0,y_{t-1})$ and $\phi_\theta(0,0,y_{t-1})$ in such a way that the above conditional expectation does not involve y_{t-1}. An appropriate guess of these four unkowns

[2] It is not difficult to see that this strategy is inapplicable with functions of two consecutive outcomes so considering functions of three consecutive outcomes is the next logical progression.

can certainly be made for the pure AR(1) case but it is arguably less immediate in multivariate specifications, especially when the support of the outcome variable is large. The particular form that we impose on $\phi_\theta(.)$ may be interpreted as a device to reduce the complexity of the problem by requiring us to solve only for $|\mathcal{Y}|$ unkowns (here two) instead of $|\mathcal{Y}|^2$ (here four). As an illustration, the proposed function $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1})$ which is null if $Y_{it} \neq 0$ satisfies:

$$E\left[\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) \mid Y_{i0} = y_0, Y_{i1}^{t-1} = y_1^{t-1}, A_i = a\right] = \frac{1}{1+e^{\gamma_0 y_{t-1}+a}}\left(\frac{e^a}{1+e^a}\phi_\theta^{0|0}(1,0,y_{t-1}) + \frac{1}{1+e^a}\phi_\theta^{0|0}(0,0,y_{t-1})\right)$$

Now, this expression makes it transparent that the only way to make the moment invariant to the value of y_{t-1} is to choose $\phi_\theta^{0|0}(1,0,y_{t-1})$ and $\phi_\theta^{0|0}(0,0,y_{t-1})$ to cancel the denominator $1+e^{\gamma_0 y_{t-1}+a}$. To achieve this, we must set: $\phi_{\theta_0}^{0|0}(1,0,y_{t-1}) = C_0 e^{\gamma_0 y_{t-1}}, \phi_{\theta_0}^{0|0}(0,0,y_{t-1}) = C_0$ for some constant $C_0 \in \mathbb{R} \setminus \{0\}$. Then,

$$E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) | Y_{i0} = y_0, Y_{i1}^{t-1} = y_1^{t-1}, A_i = a\right] = C_0 \frac{1}{1+e^a}$$

$C_0 = 1$ is the appropriate normalization to establish the connection between the function $\phi_\theta^{0|0}(.)$ and the agent specific transition probability $\pi^{0|0}(A_i)$. Of course, the same logic applies for $\phi_\theta^{1|1}(.)$ and $\pi^{1|1}(A_i)$.

Remark 1. Interestingly, Lemma 1 is a reformulation of results first shown by Kitazawa et al. (2013, 2016), Kitazawa (2022), albeit in a different way than the calculations displayed above. To see this, note that we can equivalently express the transition functions in Lemma 1 as:

$$\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}) = (e^\gamma - 1)Y_{it+1}(1-Y_{it})Y_{it-1} + (1-Y_{it})$$
$$\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}) = (e^\gamma - 1)(1-Y_{it+1})Y_{it}(1-Y_{it-1}) + Y_{it}$$

After re-arranging terms, this implies that we can write:

$$Y_{it} = (e^{\gamma_0} - 1)Y_{it+1}(1-Y_{it})Y_{it-1} + \frac{e^{A_i}}{1+e^{A_i}} + \epsilon_{it}^{0|0}, \quad E\left[\epsilon_{it}^{0|0} | Y_{i0}, Y_{i1}^{t-1}, A_i\right] = 0$$
$$Y_{it} = -(e^{\gamma_0} - 1)(1-Y_{it+1})Y_{it}(1-Y_{it-1}) + \frac{e^{\gamma_0+A_i}}{1+e^{\gamma_0+A_i}} + \epsilon_{it}^{1|1}, \quad E\left[\epsilon_{it}^{1|1} | Y_{i0}, Y_{i1}^{t-1}, A_i\right] = 0$$

These expressions coincide with the so-called *h-form* and *g-form* of Kitazawa (2022) for model (4).

We can now proceed with **Step 2)** where the ultimate goal is to leverage the transition functions to produce valid moment functions. Echoing the discussion above, one trivial way to achieve this is to consider the pairwise difference of $\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1})$ and $\phi_\theta^{k|k}(Y_{is+1}, Y_{is}, Y_{is-1})$ for any feasible $s \neq t$. This requires a minimum of four total periods of observations, counting the initial condition. Such differences automatically satisfy equation (1) by virtue of the law of iterated expectations (see the proof of Proposition 1).

Proposition 1. *In model (4) with $T \geq 3$, let*

$$\psi_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s+1}) = \phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}) - \phi_\theta^{k|k}(Y_{is+1}, Y_{is}, Y_{is-1})$$

for all $k \in \mathcal{Y}$, $t \in \{2, \dots, T-1\}$ *and* $s \in \{1, \dots, t-1\}$. *Then,*

$$E\left[\psi_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s+1})|Y_{i0}, Y_{i1}^{s-1}, A_i\right] = 0$$

4.1.2 The model with strictly exogenous regressors

We move on to the AR(1) logit model with strictly exogenous covariates characterized by equation (3). To achieve **Step 1)**, we employ the same heuristic as in the pure model and begin by looking for two moment functions $\phi_\theta^{0|0}(.)$ and $\phi_\theta^{1|1}(.)$ verifying:

$$\begin{aligned}
&\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = (1 - Y_{it})\phi_\theta^{0|0}(Y_{it+1}, 0, Y_{it-1}, X_i) \\
&\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = Y_{it}\phi_\theta^{1|1}(Y_{it+1}, 1, Y_{it-1}, X_i) \\
&E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{0|0}(A_i, X_i) \\
&E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{1|1}(A_i, X_i)
\end{aligned}$$

where $\pi_t^{k|l}(A_i, X_i)$ denotes the transition probability from state l to state k in period t:

$$\pi_t^{k|l}(A_i, X_i) = P(Y_{it+1} = k|Y_{it} = l, X_i, A_i) = P(Y_{it+1} = k|Y_{it} = l, X_{it+1}, A_i) = \frac{e^{k(\gamma_0 l + X_{it+1}'\beta_0 + A_i)}}{1 + e^{\gamma_0 l + X_{it+1}'\beta_0 + A_i}}, \quad \forall (k, l) \in \mathcal{Y}^2$$

Their expressions are provided in Lemma 2 and can be obtained following the same simple procedure described in the preceding section to derive $\phi_\theta^{0|0}(.)$ without exogenous regressors.

Lemma 2. *In model (3) with* $T \geq 2$ *and* $t \in \{1, \dots, T-1\}$, *let*

$$\begin{aligned}
&\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = (1 - Y_{it})e^{Y_{it+1}\left(\gamma Y_{it-1} - \Delta X_{it+1}'\beta\right)} \\
&\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = Y_{it}e^{(1 - Y_{it+1})\left(\gamma(1 - Y_{it-1}) + \Delta X_{it+1}'\beta\right)}
\end{aligned}$$

Then:

$$\begin{aligned}
&E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{0|0}(A_i, X_i) = \frac{1}{1 + e^{A_i + X_{it+1}'\beta_0}} \\
&E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{1|1}(A_i, X_i) = \frac{e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}{1 + e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}
\end{aligned}$$

Unlike in model (4), the covariate dependence of the transition probabilities will generally imply that the naive difference of $\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ and $\phi_\theta^{k|k}(Y_{is+1}, Y_{is}, Y_{is-1}, X_i)$ for $k \in \mathcal{Y}$ and $s \neq t$ no longer leads to valid moment functions [3]. Thus, a different technique is required in the presence of explanatory variables other than a first order lag.

The key, as foreshadowed in Section 3 is that as soon as $T \geq 3$, for periods $t \in \{2, \dots, T-1\}$, it is possible to construct transition functions other than $\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$ also associated to $\pi_t^{k|k}(A_i, X_i)$, for $k \in \mathcal{Y}$. These alternative transition functions that we denote $\zeta_\theta^{k|k}(.)$ take the form of a weighted combination of past

[3] A matching strategy in the spirit of Honoré and Kyriazidou (2000) may still be applicable but is known to lead to estimators converging at rate less than $\sqrt{N}$ for continous covariates and rules out certain regressors, e.g time dummies and time trends.

outcome $\mathbb{1}(Y_{is}=k)$, $1 \leq s < t$, and the interaction of $\mathbb{1}(Y_{is} \neq k)$ with any transition function associated to $\pi_t^{k|k}(A_i, X_i)$ having no dependence on outcomes prior to period s, e.g $\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$. This property follows from a *partial fraction decomposition* presented in Lemma 9 that exploits the structure of the model probabilities induced by the logistic assumption. It also relates to the hyperbolic transformations ideas of Kitazawa (2022). In the sequel, we will see that this insight carries over to more complex specifications like the AR(p) logit model with $p > 1$. Lemma 3 below gives the "simplest" additional transition functions that one can construct when $T \geq 3$ for the AR(1) model with exogenous regressors (the only ones when $T = 3$).

Lemma 3. *In model (3) with $T \geq 3$, for all t, s such that $T-1 \geq t > s \geq 1$, let:*

$$\begin{aligned}
\mu_s(\theta) &= \gamma Y_{is-1} + X_{is}'\beta \\
\kappa_t^{0|0}(\theta) &= X_{it+1}'\beta, \quad \kappa_t^{1|1}(\theta) = \gamma + X_{it+1}'\beta \\
\omega_{t,s}^{0|0}(\theta) &= 1 - e^{(\kappa_t^{0|0}(\theta) - \mu_s(\theta))}, \quad \omega_{t,s}^{1|1}(\theta) = 1 - e^{-(\kappa_t^{1|1}(\theta) - \mu_s(\theta))}
\end{aligned}$$

and define the moment functions:

$$\begin{aligned}
\zeta_\theta^{0|0}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i) &= (1 - Y_{is}) + \omega_{t,s}^{0|0}(\theta) Y_{is} \phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) \\
\zeta_\theta^{1|1}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i) &= Y_{is} + \omega_{t,s}^{1|1}(\theta)(1 - Y_{is}) \phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)
\end{aligned}$$

Then,

$$\begin{aligned}
E\left[\zeta_{\theta_0}^{0|0}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i) | Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] &= \pi_t^{0|0}(A_i, X_i) = \frac{1}{1 + e^{X_{it+1}'\beta_0 + A_i}} \\
E\left[\zeta_{\theta_0}^{1|1}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i) | Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] &= \pi_t^{1|1}(A_i, X_i) = \frac{e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}{1 + e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}
\end{aligned}$$

When $T \geq 4$, it turns out that we can build even more transition functions from those given in Lemma 3 by repeating the same type of logic; Corollary 3.1 provides a recursive formulation.

Corollary 3.1. *In model (3) with $T \geq 4$, for any t and ordered collection of indices s_1^J, $J \geq 2$, satisfying $T - 1 \geq t > s_1 > \ldots > s_J \geq 1$, let*

$$\begin{aligned}
\zeta_\theta^{0|0}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i) &= (1 - Y_{is_J}) + \omega_{t,s_J}^{0|0}(\theta) Y_{is_J} \zeta_\theta^{0|0}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_{J-1}-1}^{s_{J-1}}, X_i) \\
\zeta_\theta^{1|1}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i) &= Y_{is_J} + \omega_{t,s_J}^{1|1}(\theta)(1 - Y_{is_J}) \zeta_\theta^{1|1}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_{J-1}-1}^{s_{J-1}}, X_i)
\end{aligned}$$

with weights $\omega_{t,s_J}^{0|0}(\theta), \omega_{t,s_J}^{1|1}(\theta)$ defined as in Lemma 3. Then,

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

The proof is similar to that of Lemma 3 and hence omitted.

Regarding **Step 2)**, provided $T \geq 3$, the difference between any transition functions associated to the same transition probabilities in periods $t \in \{2, \ldots, T-1\}$ constitutes a valid candidate for (1). One particularly relevant set of valid moment functions is presented in Proposition 2 below.

Proposition 2. *In model (3), for all $k \in \mathcal{Y}$,*

if $T \geq 3$, for all t, s such that $T-1 \geq t > s \geq 1$, let

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

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

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

Then,

$$E\left[\psi_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] = 0$$
$$E\left[\psi_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i)|Y_{i0}, Y_{i1}^{s_J-1}, X_i, A_i\right] = 0$$

Note that this family of moment functions has cardinality $2^T - 2T$ which coincides with the number of linearly independent moments that Honoré and Weidner (2020) found numerically for the AR(1) model for any initial condition Y_{i0} [4]. To see this, notice that for fixed $(k, Y_{i0}) \in \mathcal{Y}^2$, and a given time period $t \in \{2, \ldots, T-1\}$, Proposition 2 gives a total of:

$$\sum_{l=1}^{t-1} \binom{t-1}{l} = 2^{t-1} - 1$$

valid moment functions. This follows from a simple counting argument. First, we get $\binom{t-1}{1}$ possibilities from choosing any s in $\{1, \ldots, t-1\}$ to form $\psi_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)$. To that, we must add another $\sum_{l=2}^{t-1} \binom{t-1}{l}$ possibilities from choosing all feasible sequences s_1^J with $t-1 \geq s_1 > s_2 > \ldots > s_J \geq 1$ to form $\psi_\theta^{k|k}(Y_{it-1}^{t+1}, Y_{is_1-1}^{s_1}, \ldots, Y_{is_J-1}^{s_J}, X_i)$. Summing over $t = 2, \ldots, T-1$ and multiplying by 2 to account for the two possible values for k delivers the result:

$$2 \times \sum_{t=2}^{T-1} \sum_{l=1}^{t-1} \binom{t-1}{l} = 2 \times \sum_{t=2}^{T-1} (2^{t-1} - 1) = 2 \times \left(2 \frac{1 - 2^{T-2}}{1-2} - (T-2)\right) = 2(2^{T-1} - T) = 2^T - 2T$$

For $T = 3$, the two valid moment functions produced by the model depend on two distinct sets of choice histories so are indeed linearly independent. This can be seen from their unpacked expressions in equations (13) and (14). Unfortunately, this argument does not carry over to longer panels but we have verified numerically that the linear independence property of this family continues to hold for several different values of $T \geq 4$. This suggests that our approach delivers all the *moment equality* restrictions available in the AR(1) model with T periods post initial condition [5].

[4] Dobronyi et al. (2021) proved this conjecture for the AR(1) model using a likelihood approach.

[5] This is not all the identifying content of the AR(1) specification since we know from Dobronyi et al. (2021) that the model also implies moment inequality conditions.

Remark 2. The transition functions and valid moment functions of the AR(1) model share a special symmetry property. Indeed, it is immediate that the transition functions of Lemma 2 verify

$$\phi_\theta^{0|0}(1-Y_{it+1}, 1-Y_{it}, 1-Y_{it-1}, -X_i) = \phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$$

It is then not difficult to see that this symmetry, i.e substituting Y_{it} by $(1-Y_{it})$ and X_{it} by $-X_{it}$ to obtain $\phi_\theta^{1|1}(Y_{it-1}^{t+1}, X_i)$ from $\phi_\theta^{0|0}(Y_{it-1}^{t+1}, X_i)$ transfers to the other transition functions of Lemma 3, Corollary 3.1 and ultimately to the valid moment functions of Proposition 2.

Remark 3. If $\gamma_0 = 0$, model (3) specializes to the static panel logit model of Rasch (1960) and our two-step approach is still applicable. For that specific case, Lemma 2 gives two moment functions:

$$\begin{aligned}
\phi_\theta^0(Y_{i2}, Y_{i1}, X_{i1}, X_{i2}) &= (1-Y_1)e^{-Y_{i2}\Delta X_2'\beta} \\
\phi_\theta^1(Y_{i2}, Y_{i1}, X_{i1}, X_{i2}) &= Y_{i1}e^{(1-Y_2)\Delta X_{i2}'\beta}
\end{aligned}$$

such that $E\left[\phi_{\theta_0}^0(Y_{i2}, Y_{i1}, X_{i1}, X_{i2})|X_i, A_i\right] = \frac{1}{1+e^{X_{i2}'\beta_0+A_i}}$ and $E\left[\phi_{\theta_0}^1(Y_{i2}, Y_{i1}, X_{i1}, X_{i2})|X_i, A_i\right] = \frac{e^{X_{i2}'\beta_0+A_i}}{1+e^{X_{i2}'\beta_0+A_i}}$. It follows that a valid moment function with two periods of observation is

$$\begin{aligned}
\psi_\theta(Y_{i2}, Y_{i1}, X_{i1}, X_{i2}) &= \phi_\theta^1(Y_{i2}, Y_{i1}, X_{i1}, X_{i2}) - (1-\phi_\theta^0(Y_{i2}, Y_{i1}, X_{i1}, X_{i2})) \\
&= (1-e^{-\Delta X_{i2}'\beta})\left(Y_{i1}(1-Y_{i2})e^{\Delta X_{i2}'\beta} - (1-Y_{i1})Y_{i2}\right)
\end{aligned}$$

which is proportional to the score of the conditional likelihood based on the sufficient statistic $Y_{i1}+Y_{i2}$ (see for example Bonhomme (2012), equation (21)).

Remark 4. The methodology developed in this section will apply more generally to models where A_i is additively separable from the other explanatory variables. That is,

$$Y_{it} = \mathbb{1}\left\{g(Y_{it-1}, X_{it}, \theta_0) + A_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, \ldots, T$$

where $g(.)$ is known up to the common parameter θ_0. This means that interactions between Y_{it-1} and X_{it} can be included as regressors without any fundamental changes.

4.2 Simple application to an AR(1) logit with a time dummy

The AR(1) model with time dummies is one of the prototypical examples where the matching strategy of Honoré and Kyriazidou (2000) fails to apply. In this type of situations, the moment conditions of Proposition 2 become instrumental and can even point identify the model parameters under certain conditions. To illustrate this point concretely, consider specification (3) with $X_{it} = \mathbb{1}\{t=2\}$, for all $t \in \{1, \ldots, T\}$; reflecting an instance where the econometrician may want to control for the effect of a particular period due perhaps to a one-off policy. This toy example has the purposeful feature that X_{i2} is different from X_{i3} so as to invalidate the conditional likelihood approach.

With $T=3$, Proposition 2 delivers two valid moment functions per initial condition. For $Y_{i0}=0$, they read :

$$\psi_\theta^{0|0}(Y_{i1}^3, 0, X_i) = (e^{\beta}-1)(1-Y_{i1})(1-Y_{i2})Y_{i3} + e^{\beta+\gamma}Y_{i1}(1-Y_{i2})Y_{i3} + Y_{i1}(1-Y_{i2})(1-Y_{i3}) - (1-Y_{i1})Y_{i2} \tag{5}$$

$$\psi_\theta^{1|1}(Y_{i1}^3, 0, X_i) = (e^{-\beta}-1)Y_{i1}Y_{i2}(1-Y_{i3}) + e^{-\beta}(1-Y_{i1})Y_{i2}(1-Y_{i3}) + e^{-\gamma}(1-Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1-Y_{i2}) \tag{6}$$

whereas their symmetric counterpart for $Y_{i0}=1$ are [6]:

$$\psi_\theta^{0|0}(Y_{i1}^3, 1, X_i) = (e^{\beta}-1)(1-Y_{i1})(1-Y_{i2})Y_{i3} + e^{\beta}Y_{i1}(1-Y_{i2})Y_{i3} + e^{-\gamma}Y_{i1}(1-Y_{i2})(1-Y_{i3}) - (1-Y_{i1})Y_{i2} \tag{7}$$

$$\psi_\theta^{1|1}(Y_{i1}^3, 1, X_i) = (e^{-\beta}-1)Y_{i1}Y_{i2}(1-Y_{i3}) + e^{-\beta+\gamma}(1-Y_{i1})Y_{i2}(1-Y_{i3}) + (1-Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1-Y_{i2}) \tag{8}$$

Taken separately, (5)-(6) and (7)-(8) are in general not enough to point identify (γ_0, β_0) due to the possible presence of a false root [7]. This is shown in Figure 1 which plots the set of parameter values for which the moment restrictions for $Y_{i0}=0$ (left panel) and $Y_{i0}=1$ (right panel) hold in an example data generating process. Jointly however, they are sufficient to pin down the true parameter as we now demonstrate. It suffices to take (appropriate) conditional expectations of (6) and (7) and recast the resulting moment restrictions as:

$$\begin{aligned}
e^{-\gamma_0} &= \frac{P_0(1,1,0)+P_0(1,0)}{P_0(0,1,1)} - \frac{P_0(1,1,0)+P_0(0,1,0)}{P_0(0,1,1)}e^{-\beta_0}\\
e^{-\gamma_0} &= \frac{P_1(0,0,1)+P_1(0,1)}{P_1(1,0,0)} - \frac{P_1(0,0,1)+P_1(1,0,1)}{P_1(1,0,0)}e^{\beta_0}
\end{aligned}$$

In the above expressions, we use the shorthands $P_{y_0}(y_1,\dots,y_s)$ with $s>1$ to denote the probability of choice history $(y_1,\dots,y_s)$ given the initial condition y_0. Substracting off the second equation from the first highlights that β_0 is a root of the exponential polynomial

$$Q(\beta) = \left(\frac{P_0(1,1,0)+P_0(1,0)}{P_0(0,1,1)} - \frac{P_1(0,0,1)+P_1(0,1)}{P_1(1,0,0)}\right) - \frac{P_0(1,1,0)+P_0(0,1,0)}{P_0(0,1,1)}e^{-\beta} + \frac{P_1(0,0,1)+P_1(1,0,1)}{P_1(1,0,0)}e^{\beta}$$

It is readily verified that $Q(\beta)$ is strictly increasing on the real line with $\lim_{\beta\to-\infty} Q(\beta) = -\infty$ and $\lim_{\beta\to+\infty} Q(\beta) = +\infty$. Thus, β_0 must be its unique root. Since the informative moment conditions implied by (5) and (8) were unexploited in this argument, we further conclude that the model parameters are overidentified with $T=3$ if there is variablity in the initial condition.

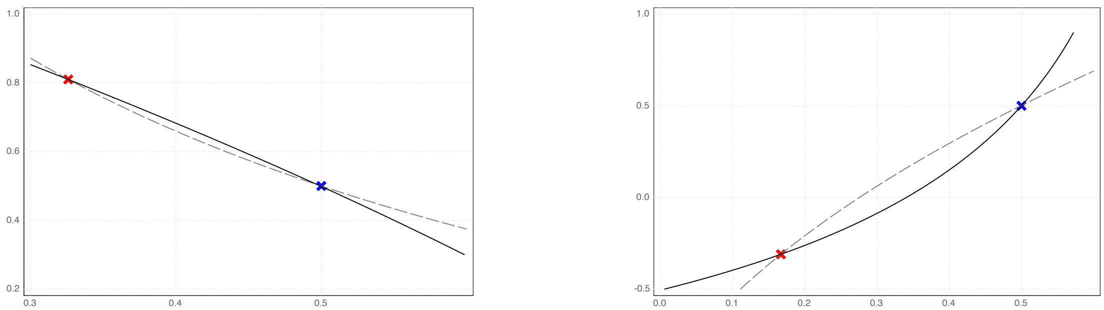
Hahn (2001) studied a more complex version with two time dummies and $T=3$ wherein point identification based solely on moment equality conditions is not guaranteed even if both $Y_{i0}=0$ and $Y_{i0}=1$

[6] We refer the reader to equations (14) and (13) in Appendix Section E to see how these expressions can be obtained.

[7] It is possible to show that there are at most two roots for any initial condition and we leave this exercise to the reader.

Figure 1: Roots of the two moment conditions for $T = 3$

(a) For $Y_{i0} = 0$ (b) For $Y_{i0} = 1$



Notes: The DGP for this figure sets $\beta_0 = \gamma_0 = 0.5$ with $Y_{i0} \sim^{iid} Bernouilli(\frac{1}{2})$, $A_i \sim^{iid} \mathcal{N}(0, 1)$, $X_{it} = \mathbb{1}\{t = 2\}$ and ϵ_{it} iid standard logistic. The x-axis and y-axis represent the values for β and γ respectively. The solid black lines in panel (a), respectively (b) depict the set of parameter values for which the moment conditions based on (5), respectively (6) hold. Similarly, the dashed grey lines in panel (a), respectively (b) depict the set of parameter values for which the moment conditions based on (7), respectively (8) hold. The blue cross indicates the correct root, i.e (0.5,0.5) while the red cross indicates the false root.

are observed (Dobronyi et al. (2021)). In that case, for some data generating processes, harnessing moment inequality conditions implied by the model likelihood may help restore point identification as shown by Dobronyi et al. (2021). The goal of our more stylized example is to plainly illustrate the usefulness of moment conditions in a situation where conditional likelihood fails within the confines of the AR(1) logit model.

4.3 Average Marginal Effects in the AR(1) logit model

In discrete choice settings, interest often centers on functionals of unobserved heterogeneity rather than on the value of the model parameters per se. One such functional of interest from a policy perspective is the average marginal effects (AMEs) which capture mean response to a counterfactual change in past outcomes. In the context of the AR(1) model, Aguirregabiria and Carro (2021) established that AMEs are point identified given sufficient observations. We revisit their conclusions through the lens of our methodology and show how these key quantities can trivially be computed from the model transition functions.

First, for the pure model described by equation (4), the average transition probability from state l to state k for the initial condition y_0 are defined as:

$$\Pi^{k|l}(y_0) = E\left[\pi^{k|l}(A_i)|Y_{i0} = y_0\right] = \int \pi^{k|l}(a)p(a|y_0)da, \quad \forall (k,l) \in \mathcal{Y}^2$$

where $p(a|y_0)$ denotes the marginal probability of the fixed effect A given y_0. The corresponding AME is the following contrast of average transition probabilities:

$$AME(y_0) = \Pi^{1|1}(y_0) - \Pi^{1|0}(y_0) = \Pi^{1|1}(y_0) - (1 - \Pi^{0|0}(y_0))$$

It is interpreted as the population average causal effect on Y_{it+1} of a change from 0 to 1 of Y_{it} given y_0. We now make the useful observation that so long as $T \geq 2$, for any $t \geq 1$ we have by Lemma 1:

$$\Pi^{0|0}(y_0) = E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1})|Y_{i0} = y_0\right]$$
$$\Pi^{1|1}(y_0) = E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1})|Y_{i0} = y_0\right]$$

Given that γ_0 is point identified for $T \geq 3$, either by the moment conditions of Corollary 1 (see also Kitazawa (2022)) or by the traditional conditional likelihood approach (Chamberlain (1985), Honoré and Kyriazidou (2000), Magnac (2000)), these formulas imply that the AME and average transition probabilities will also be point identified in that case.

A similar observation holds in the model with exogenous regressors given by the law of motion (3). In that more general setting, one may consider the average transition probability from state l to state k in period t for a subpopulation of individuals with covariate $x = (x_1, \dots, x_T)$. Formally,

$$\Pi_t^{k|l}(y_0, x) = E\left[\pi_t^{k|l}(X_i, A_i)\,|\,Y_{i0} = y_0, X_i = x\right] = \int \pi_t^{k|l}(x, a)p(a|y_0, x)da, \quad \forall (k, l) \in \mathcal{Y}^2, \quad \forall x \in \mathbb{R}^{TK_x}$$

where this time $p(a|y_0, x)$ denotes the marginal probability of the fixed effect A given y_0 and regressor x. Then, the corresponding AMEs are simply:

$$AME_t(y_0, x) = \Pi_t^{1|1}(y_0, x) - \Pi_t^{1|0}(y_0, x) = \Pi_t^{1|1}(y_0, x) - (1 - \Pi_t^{0|0}(y_0, x))$$

It follows from Lemma 2 that for $T \geq 2$ and $t \geq 1$:

$$\Pi_t^{0|0}(y_0, x) = E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)\,|\,Y_{i0} = y_0, X_i = x\right]$$
$$\Pi_t^{1|1}(y_0, x) = E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)\,|\,Y_{i0} = y_0, X_i = x\right]$$

and if four consecutive periods or more are available to the econometrician, these functionals can also be computed as expectations of the transition functions provided in Lemma 3 and Corollary 3.1. In Appendix Section E, we show that our moment conditions coincide exactly with those of Honoré and Weidner (2020) for $T = 3$. As a result, we know without further work that if $T \geq 3$ the moment functions of Proposition 2 identify the common parameters $\theta_0 = (\gamma_0, \beta_0') \in \mathbb{R}^{K_x+1}$ if X_{it} consist of continuous and or discrete regressors. In turn, average transition probabilities and AMEs are point identified under the same conditions. Remark 4 indicates that this result may carry over to extensions of the baseline model with regressors of the form $Z_{it} = f(Y_{it-1}, X_{it})$.

4.4 Connections to other works on the AR(1) logit model

As highlighted in the pure model, there is an insightful parallel between our methodology and that of Kitazawa (2022). This is equally true in the AR(1) model with strictly exogenous explanatory variables. Indeed, after some algebraic manipulation, we can re-express the transition functions of Lemma 2 as:

$$
\begin{aligned}
\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) &= 1 - Y_{it} - (1 - Y_{it})Y_{it+1} + (1 - Y_{it})Y_{it+1}e^{-\Delta X'_{it+1}\beta} + \delta Y_{it-1}(1 - Y_{it+1})Y_{it+1}e^{-\Delta X'_{it+1}\beta} \\
\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) &= Y_{it}Y_{it+1} + Y_{it}(1 - Y_{it+1})e^{\Delta X'_{it+1}\beta} + \delta(1 - Y_{it-1})Y_{it}(1 - Y_{it+1})
\end{aligned}
$$

where $\delta = (e^\gamma - 1)$. Therefore, the moment conditions of Lemma 2 imply that we can write:

$$
\begin{aligned}
Y_{it} + (1 - Y_{it})Y_{it+1} - (1 - Y_{it})Y_{it+1}e^{-\Delta X'_{it+1}\beta_0} - \delta_0 Y_{it-1}(1 - Y_{it+1})Y_{it+1})e^{-\Delta X'_{it+1}\beta_0} &= \frac{e^{X'_{it+1}\beta_0 + A_i}}{1 + e^{X'_{it+1}\beta_0 + A_i}} + \epsilon_{it}^{0|0} \\
Y_{it}Y_{it+1} + Y_{it}(1 - Y_{it+1})e^{\Delta X'_{it+1}\beta_0} + \delta_0(1 - Y_{it-1})Y_{it}(1 - Y_{it+1}) &= \frac{e^{\gamma_0 + X'_{it+1}\beta_0 + A_i}}{1 + e^{\gamma_0 + X'_{it+1}\beta_0 + A_i}} + \epsilon_{it}^{1|1}
\end{aligned}
$$

where $E\left[\epsilon_{it}^{0|0}|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = 0$ and $E\left[\epsilon_{it}^{1|1}|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = 0$. These expressions are the *h-form* and *g-form* of Kitazawa (2022) for model (3) and were originally obtained through an ingenious usage of the mathematical properties of the hyperbolic tangent function. The evident connection between the transition functions and the *h-form* and *g-form* offers an interesting new perspective on the transformation approach of Kitazawa (2022) for the AR(1) model.

If we further define

$$
\begin{aligned}
U_{it} &= Y_{it} + (1 - Y_{it})Y_{it+1} - (1 - Y_{it})Y_{it+1}e^{-\Delta X'_{it+1}\beta} - \delta Y_{it-1}(1 - Y_{it+1})Y_{it+1})e^{-\Delta X'_{it+1}\beta} \\
\Upsilon_{it} &= Y_{it}Y_{it+1} + Y_{it}(1 - Y_{it+1})e^{\Delta X'_{it+1}\beta} + \delta(1 - Y_{it-1})Y_{it}(1 - Y_{it+1})
\end{aligned}
$$

the two moment functions of Kitazawa (2022) for the AR(1) model write

$$
\begin{aligned}
\hbar U_{it} &= U_{it} - Y_{it-1} - \tanh\left(\frac{-\gamma Y_{it-2} + (\Delta X_{it} + \Delta X_{it+1})'\beta}{2}\right)(U_{it} + Y_{it-1} - 2U_{it}Y_{it-1}) \\
\hbar \Upsilon_{it} &= \Upsilon_{it} - Y_{it-1} - \tanh\left(\frac{\gamma(1 - Y_{it-2}) + (\Delta X_{it} + \Delta X_{it+1})'\beta}{2}\right)(\Upsilon_{it} + Y_{it-1} - 2\Upsilon_{it}Y_{it-1})
\end{aligned}
$$

which can be formulated in terms of our own moment functions as

$$
\begin{aligned}
\hbar U_{it} &= -\frac{2}{2 - \omega_{t,t-1}^{0|0}(\theta)}\psi_\theta^{0|0}(Y_{it-1}^{t+1}, Y_{it-2}^{t-1}, X_i) \\
\hbar \Upsilon_{it} &= \frac{2}{2 - \omega_{t,t-1}^{1|1}(\theta)}\psi_\theta^{1|1}(Y_{it-1}^{t+1}, Y_{it-2}^{t-1}, X_i)
\end{aligned}
$$

Appendix Section E provides detailed derivations for the mapping between our two approaches. This last result indicates that our moment conditions essentially match those of Kitazawa (2022) when $T = 3$. It is certain however that for $T \geq 4$, Proposition 2 imply that there are further identifying moments than those based solely on $\hbar U_{it}$ and $\hbar \Upsilon_{it}$ for the AR(1) model. Interestingly, as mentioned previously, it turns out as we show in Appendix Section E that our moment functions coincide exactly with those derived by Honoré and Weidner (2020) for the special case $T = 3$.

To the best of our knowledge, besides the AR(1) model and a few specific examples, the structure of moment conditions in models with arbitrary lag order is not fully understood in the literature. Building on

Bonhomme (2012), Honoré and Weidner (2020) propose moment functions for the AR(2) model up to $T = 4$ and the AR(3) model with $T = 5$ but no results are offered beyond these special instances. Yet, we believe this is of general interest not only to better understand the properties of DFEL models but also for practical modelling and estimation purposes. For example, Card and Hyslop (2005) argue in favor of using higher order logit specifications to model a control group in the context of a welfare experiment. Relatedly, there are few results available for multivariate fixed effect models and existing methods developed for the scalar case are likely to be difficult to adapt due to computational barriers. In the remaining sections, we show that our two-step approach addresses these issues by providing closed form expressions for the moment equality conditions of these more complex models.

4.5 Moment conditions for the AR(p) logit model, $p > 1$

Allowing for more than one lag is often desirable in empirical work to model persistent stochastic processes and to better fit the data (e.g Magnac (2000) on labour market histories, Chay et al. (1999) and Card and Hyslop (2005) on welfare recipiency). To this end, we now discuss how to extend our identification scheme to general univariate autoregressive models. We consider

$$Y_{it} = \mathbb{1}\left\{\sum_{r=1}^{p} \gamma_{0r} Y_{it-r} + X_{it}'\beta_0 + A_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, \dots, T \tag{9}$$

for known autoregressive order $p > 1$ and vector of initial values $Y_i^0 = (Y_{i-(p-1)}, \dots, Y_{i-1}, Y_{i0})' \in \mathcal{Y}^p$, with $A_i \in \mathbb{R}$.

The corresponding one-period ahead transition probabilities are given by:

$$\pi_t^{k|l_1^p}(A_i, X_i) = P(Y_{it+1} = k | Y_{it} = l_1, \dots, Y_{it-(p-1)} = l_p, X_i, A_i) = \frac{e^{k(\sum_{l=1}^{p} \gamma_{0r} l_r + X_{it+1}'\beta_0 + A_i)}}{1 + e^{\sum_{r=1}^{p} \gamma_{0r} l_r + X_{it+1}'\beta_0 + A_i}}, \quad \forall (k, l_1, \dots, l_p) \in \mathcal{Y}^{p+1}$$

The flexibility granted by model (9) relative to the baseline AR(1) necessarily comes with further technical challenges but the overall methodology to derive moment conditions remains unchanged. To reiterate, **Step 1)** concerns the computation of the model transition functions and **Step 2)** consists in differencing these transition functions to output moments functions verifying (1). In fact, notice that upon completion of **Step 1)**, **Step 2)** in the AR(p) model with $p > 1$ poses no further challenges than in the AR(1) model. This is because the transition probabilities take the same functional form: a logistic transformation of a linear index composed of common parameters, the regressors and the fixed effect only. Hence, the same differencing strategy applies to produce valid moment functions. Complications only arise in **Step 1)** owing to the presence of further lagged outcomes relative to specification (3). This is addressed in Theorem 1 which provides an algorithm to compute a set of transition functions for arbitrary lag order greater than one.

Theorem 1. *In model (9) with $T \geq p+1$, for all $t \in \{p, \dots, T-1\}$ and $y_1^p \in \mathcal{Y}^p$, let*

$$
\begin{aligned}
&k_t^{y_1|y_1^p}(\theta) = \sum_{r=1}^{p} \gamma_r y_r + X'_{it+1}\beta \\
&k_t^{y_1|y_1^{k+1}}(\theta) = \sum_{r=1}^{k+1} \gamma_r y_r + \sum_{r=k+2}^{p} \gamma_r Y_{it-(r-1)} + X'_{it+1}\beta, \quad k = 1, \dots, p-2, \ \textit{if } p > 2 \\
&u_{t-k}(\theta) = \sum_{r=1}^{p} \gamma_r Y_{it-(r+k)} + X'_{it-k}\beta, \quad k = 1, \dots, p-1 \\
&w_t^{y_1|y_1^{k+1}}(\theta) = \left[1 - e^{(k_t^{y_1|y_1^{k+1}}(\theta) - u_{t-k}(\theta))}\right]^{y_{k+1}} \left[1 - e^{-(k_t^{y_1|y_1^{k+1}}(\theta) - u_{t-k}(\theta))}\right]^{1-y_{k+1}}, \quad k = 1, \dots, p-1
\end{aligned}
$$

and

$$
\begin{aligned}
&\phi_\theta^{y_1|y_1^{k+1}}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i) = \\
&\left[(1 - Y_{it-k}) + w_t^{y_1|y_1^{k+1}}(\theta)\phi_\theta^{y_1|y_1^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)Y_{it-k}\right]^{(1-y_1)y_{k+1}} \times \\
&\left[1 - Y_{it-k} - w_t^{y_1|y_1^{k+1}}(\theta)\left(1 - \phi_\theta^{y_1|y_1^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)\right)(1 - Y_{it-k})\right]^{(1-y_1)(1-y_{k+1})} \times \\
&\left[Y_{it-k} + w_t^{y_1|y_1^{k+1}}(\theta)\phi_\theta^{y_1|y_1^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)(1 - Y_{it-k})\right]^{y_1(1-y_{k+1})} \times \\
&\left[1 - (1 - Y_{it-k}) - w_t^{y_1|y_1^{k+1}}(\theta)\left(1 - \phi_\theta^{y_1|y_1^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)\right)Y_{it-k}\right]^{y_1 y_{k+1}}, \quad k = 1, \dots, p-1
\end{aligned}
$$

where

$$
\begin{aligned}
\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i) &= (1 - Y_{it})e^{Y_{it+1}(\gamma_1 Y_{it-1} - \sum_{l=2}^{p} \gamma_l \Delta Y_{it+1-l} - \Delta X'_{it+1}\beta)} \\
\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i) &= Y_{it}e^{(1-Y_{it+1})\left(\gamma_1(1-Y_{it-1}) + \sum_{l=2}^{p} \gamma_l \Delta Y_{it+1-l} + \Delta X'_{it+1}\beta\right)}
\end{aligned}
$$

Then,

$$
\begin{aligned}
&E\left[\phi_{\theta_0}^{y_1|y_1^p}(Y_{it+1}, Y_{it}, Y_{it-(2p-1)}^{t-1}, X_i) \,|\, Y_i^0, Y_{i1}^{t-p}, X_i, A_i\right] = \pi_t^{y_1|y_1^p}(A_i, X_i) \\
&E\left[\phi_{\theta_0}^{y_1|y_1^{k+1}}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i) \,|\, Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right] = \pi_t^{y_1|y_1^{k+1}, Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i), \quad k = 0, \dots, p-2
\end{aligned}
$$

Amongst other things, the theorem shows that closed form expressions of the model transition functions in periods $t \in \{p, \dots, T-1\}$ exist so long as $T \geq p+1$.

It is insightful to discuss in details the simple example of the AR(2) model to understand the intuition behind Theorem 1 and clarify how it addresses the issue of higher order lags. To fix ideas, suppose $T \geq 3$ and that we are interested in finding a transition function associated to the transition probability

$$
\pi_t^{0|0,1}(A_i, X_i) = \frac{1}{1 + e^{\gamma_{02} + X'_{it+1}\beta_0 + A_i}}
$$

for some $t \in \{2, \ldots, T-1\}$. The starting point of Theorem 1 is to reason by analogy with the AR(1) model and search for a moment function $\phi_\theta^{0|0}(.)$ of the form:

$$\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, Y_{it-2}, X_i) = (1 - Y_{it})\phi_\theta^{0|0}(Y_{it+1}, 0, Y_{it-1}, Y_{it-2}, X_i)$$

in the hope that some transition function associated to $\pi_t^{0|0,1}(A_i, X_i)$ has this exact structure. Notice the dependence on four consecutive outcomes, $Y_{it+1}, Y_{it}, Y_{it-1}, Y_{it-2}$, instead of three which reflects the intention of accounting for one additional lagged outcome in the AR(2) model. Calculations analogous to those for the pure AR(1) in Section 4 quickly reveal that the problem cannot be solved directly in this way but that we can nevertheless select $\phi_\theta^{0|0}(.)$ such that:

$$E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{0|0,Y_{it-1}}(A_i, X_i)$$

It requires setting

$$\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i) = (1 - Y_{it})e^{Y_{it+1}(\gamma_1 Y_{it-1} - \gamma_2 \Delta Y_{it-1} - \Delta X'_{it+1}\beta)}$$

We have reached an intermediate step since $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i)$ maps to $\pi_t^{0|0,Y_{it-1}}(A_i, X_i)$ which differs from $\pi_t^{0|0,1}(A_i, X_i)$ due to its dependence on the random variable Y_{it-1}. To make further progress, one would intuitively need to "set" Y_{it-1} to 1 so that the two transition probabilities coincide. Naturally, we cannot directly manipulate Y_{it-1} as it is predetermined but we can make this idea more concrete by interacting $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i)$ with Y_{it-1}. Computing the conditional expectation of this product and using iterated expectations, one gets

$$\begin{aligned} E\left[Y_{it-1}\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-2}, X_i, A_i\right] &= E\left[Y_{it-1}\pi_t^{0|0,1}(A_i, X_i)|Y_i^0, Y_{i1}^{t-2}, X_i, A_i\right] \\ &= \frac{1}{1 + e^{\gamma_{02} + X'_{it+1}\beta + A_i}} \frac{e^{\gamma_{01} Y_{it-2} + \gamma_{02} Y_{it-3} + X'_{it-1}\beta_0 + A_i}}{1 + e^{\gamma_{01} Y_{it-2} + \gamma_{02} Y_{it-3} + X'_{it-1}\beta_0 + A_i}} \end{aligned}$$

At this point, we recognize a product of logistic indices similar to those encountered in the construction of sequences of transition functions for the AR(1) model (see for example Lemma 3). Hence, appealing to the same partial fraction decomposition in Appendix Lemma 9, we infer that a transition function associated to $\pi_t^{0|0,1}(A_i, X_i)$ exists in the form of a weighted average of $(1 - Y_{it-1})$ and the interaction of $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-2}^{t-1}, X_i)$ with Y_{it-1}. Specifically, it reads:

$$\phi_\theta^{0|0,1}(Y_{it+1}, Y_{it}, Y_{it-3}^{t-1}, X_i) = (1 - Y_{it-1}) + (1 - e^{\gamma_2 + X'_{it+1}\beta - (\gamma_1 Y_{it-2} + \gamma_2 Y_{it-3} + X'_{it-1}\beta)})\phi_\theta^{0|0}(Y_{it-2}^{t+1}, X_i)Y_{it-1}$$

Theorem 1 generalizes this procedure to the other transition probabilities, viz. $\pi_t^{0|0,0}(A_i, X_i), \pi_t^{1|1,0}(A_i, X_i)$, $\pi_t^{1|1,1}(A_i, X_i)$, and to models with a higher lag order. For example, if $p = 3$, the algorithm goes through 3 steps to compute the transition function associated to say, $\pi_t^{0|0,1,0}(A_i, X_i)$. The first step finds the moment function associated to $\pi_t^{0|0,Y_{it-1},Y_{it-2}}(A_i, X_i)$. The second step eliminates the dependence on Y_{it-1} from the first step and yields the moment function associated to $\pi_t^{0|0,1,Y_{it-2}}(A_i, X_i)$. Lastly, the third step eliminates the dependence on Y_{it-2} from the second step and produces the target transition function. There will be p steps of the same nature when the autoregressive order is p, with $p > 1$.

As advertised, **Step 2)** is now analogous to the AR(1) case since the transition probabilities keep an identical structure. As soon as $T \geq p+2$, for periods $t \in \{p+1, \ldots, T-1\}$, we can construct transition functions other than $\phi_\theta^{y_1|y_1^p}(Y_{it+1}, Y_{it}, Y_{it-(2p-1)}^{t-1}, X_i)$ also associated to $\pi_t^{y_1|y_1^p}(A_i, X_i)$, for $y_1^p \in \mathcal{Y}^p$. These new transition functions that we denote $\zeta_\theta^{y_1|y_1^p}(.)$ take the form of a weighted combination of past outcome $\mathbb{1}(Y_{is} = y_1)$, $s \in \{1, \ldots, t-p\}$ and the interaction of $\mathbb{1}(Y_{is} \neq y_1)$ with any transition function whose conditiong set encompasses Y_{is} for it to map to $\pi_t^{y_1|y_1^p}(A_i, X_i)$. The simplest examples which are also the only ones available when $T = p+2$, are given in Lemma 4.

Lemma 4. *In model (9) with $T \geq p+2$, for all $t \in \{p+1, \ldots, T-1\}$, $s \in \{1, \ldots, t-p\}$ and $y_1^p \in \mathcal{Y}^p$, let*

$$\mu_s(\theta) = \sum_{r=1}^{p} \gamma_{0r} Y_{is-r} + X_{is}'\beta$$
$$\kappa_t^{y_1|y_1^p}(\theta) = \sum_{r=1}^{p} \gamma_{0r} y_r + X_{it+1}'\beta$$
$$\omega_{t,s}^{y_1|y_1^p}(\theta) = \left[1 - e^{(\kappa_t^{y_1|y_1^p}(\theta) - \mu_s(\theta))}\right]^{1-y_1} \left[1 - e^{-(\kappa_t^{y_1|y_1^p}(\theta) - \mu_s(\theta))}\right]^{y_1}$$

and define the moment functions:

$$\zeta_\theta^{0|0,y_2^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i) = (1 - Y_{is}) + \omega_{t,s}^{0|0,y_2^p}(\theta) Y_{is} \phi_\theta^{0|0,y_2^p}(Y_{it+1}, Y_{it}, Y_{it-(2p-1)}^{t-1}, X_i)$$
$$\zeta_\theta^{1|1,y_2^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i) = Y_{is} + \omega_{t,s}^{1|1,y_2^p}(\theta)(1 - Y_{is}) \phi_\theta^{1|1,y_2^p}(Y_{it+1}, Y_{it}, Y_{it-(2p-1)}^{t-1}, X_i)$$

Then,

$$E\left[\zeta_{\theta_0}^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i) | Y_i^0, Y_{i1}^{s-1}, X_i, A_i\right] = \pi_t^{y_1|y_1^p}(A_i, X_i)$$

Unsurprisingly, as in the AR(1) case, it becomes possible to construct iteratively more transition functions from those given in Lemma 4 when at least $T = p+3$ periods are observed post initial condition. They are given in Corollary 4.1 below.

Corollary 4.1. *In model (9) with $T \geq p+3$, 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*

$$\zeta_\theta^{0|0,y_2^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i) = (1 - Y_{is_J}) + \omega_{t,s_J}^{0|0,y_2^p}(\theta) Y_{is_J} \zeta_\theta^{0|0,y_2^p}(Y_{it-1}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_{J-1}-p}^{s_{J-1}}, X_i)$$
$$\zeta_\theta^{1|1,y_2^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i) = Y_{is_j} + \omega_{t,s_J}^{1|1,y_2^p}(\theta)(1 - Y_{is_J}) \zeta_\theta^{1|1,y_2^p}(Y_{it-1}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_{J-1}-p}^{s_{J-1}}, X_i)$$

with weights $\omega_{t,s_J}^{y_1|y_1^p}(\theta)$ defined as in Lemma 4. Then,

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

The proof follows the same logic as that of Lemma 4 and is thus omitted.

For **Step 2)**, provided that $T \geq p+2$, 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 3 hereinbelow presents one set of valid moment functions that generalize those obtained previously for the one lag case.

Proposition 3. *In model (9)*
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,

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

This family of moment functions features precisely $2^T - (T+1-p)2^p$ distinct elements for any initial condition. Indeed, fix Y_i^0 and a p-vector $y_1^p \in \{0,1\}^p$. Then, for a given time period $t \in \{p+1, \ldots, T-1\}$, there are $\binom{t-p}{1}$ moments of the form $\psi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is-p}^{s}, X_i)$ corresponding to choices of $s \in \{1, \ldots, t-p\}$. Moreover, by choosing any feasible sequence s_1^J, $J \geq 2$, verifying $t-p \geq s_1 > \ldots > s_J \geq 1$ we produce another $\sum_{l=2}^{t-p} \binom{t-p}{l}$ moment functions of the form $\psi_\theta^{y_1|y_1^p}(Y_{it-(2p-1)}^{t+1}, Y_{is_1-p}^{s_1}, \ldots, Y_{is_J-p}^{s_J}, X_i)$. In total, for period t, we count :

$$\sum_{l=1}^{t-p} \binom{t-p}{l} = 2^{t-p} - 1$$

valid moments. Now, summing over all possible values for $t \in \{p+1, \ldots, T-1\}$ and multiplying by the number of distinct values for y_1^p, namely 2^p, we get:

$$2^p \sum_{t=p+1}^{T-1} \sum_{l=1}^{t-p} \binom{t-p}{l} = 2^p \sum_{t=p+1}^{T-1} (2^{t-p} - 1) = 2^p \left(2\frac{1-2^{T-p-1}}{1-2} - (T-p-1)\right) = 2^T - (T+1-p)2^p$$

Numerical experimentation for various values of T in the AR(1) and AR(2) cases suggest that the moment functions of Proposition 3 are effectively linearly independent. Therefore, collectively, these results seem to corroborate the conjecture of Honoré and Weidner (2020) that the AR(p) model with $T \geq p+2$ periods induces a total of $2^T - (T+1-p)2^p$ linearly independent moment conditions. This also suggests that functional differencing at least in panel data logit models can be broken down into a series of equivalent simpler subproblems period by period that find all moment equality restrictions. As we illustrate, such an approach can be particularly fruitful to tackle more complex models where an analysis of the full likelihood, a high dimensional object, may prove difficult.

Remark 5. Even though the exposition centered on model (9), our methodology applies more broadly to models of the form

$$Y_{it} = \mathbb{1}\left\{g(Y_{it-1}, \ldots, Y_{it-p}, X_{it}, \theta_0) + A_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, \ldots, T$$

where the lag order $p > 1$ is known and $g(.)$ is known up to the finite dimensional parameter θ_0. We can thus incorporate interaction effects within our framework which are often of interest in applied work. For instance, Card and Hyslop (2005) model welfare participation as a random effect AR(2) logit process with an interaction term between lagged outcomes. It takes the form

$$Y_{it} = \mathbb{1}\left\{\gamma_{01} Y_{it-1} + \gamma_{02} Y_{it-2} + \delta_0 Y_{it-1} Y_{it-2} + X_{it}'\beta_0 + A_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, \ldots, T$$

where A_i is the random effect following either a normal distribution or a discrete distribution with few support points. In this case, minor modifications of the results in this section will deliver moment conditions for $\theta_0 = (\gamma_{01}, \gamma_{01}, \delta_0, \beta_0')$ that are immune to misspecifications of individual unobserved heterogeneity.

4.6 Impossibility of moment restrictions in AR(p) logit models with $T \leq p + 1$

Our results so far show how to construct moment restrictions on common parameters when at least $T = p+2$ observations after $t = 0$ are available. In Theorem 2, we establish that this is effectively the minimum number of time periods required for the existence of valid moment functions in AR(p) models.

Theorem 2. *Consider model (9) with $p \geq 1$ and initial condition $y^0 \in \mathcal{Y}^p$. Suppose that $\gamma_{0r} \neq 0$ for all $r \in \{1, \ldots, p\}$ and $\beta_0 \neq 0$ with $x_t'\beta_0 \neq x_s'\beta_0$ for all $t \neq s$. Then, for any $T \leq p+1$, there are no valid moment functions in the sense of equation (1).*

Honoré and Weidner (2020) showed a simpler impossibility result for the special case of the AR(1) with $T = 2$. To prove Theorem 2, we start from the observation that equation (1) for a given T, initial condition y^0 and regressors x can be written equivalently as:

$$\sum_{(y_1,\ldots,y_T) \in \mathcal{Y}^T} \psi_{\theta_0}(y_1^T, y^0, x) P(Y_{i1} = y_1, \ldots, Y_{iT} = y_T | Y_i^0 = y^0, X_i = x, A_i = a) = 0, \quad \forall a \in \mathbb{R}$$

This formulation clarifies that the existence of a valid moment function for a given T is equivalent to the existence of a linear relationship between the conditional probabilities of all choice histories of length T, $y_1^T = (y_1, \ldots, y_T) \in \mathcal{Y}^T$, given (y^0, x, a) when viewed as functions of a. We then show using an induction argument and elementary properties of polynomials that such a linear dependence does not exist when $T \leq p + 1$. An immediate corollary is that the transition functions in period $t = p$ whose expression are provided in Theorem 1 are unique under such conditions.

The conditions of the theorem are the most general insofar as all model parameters are relevant and the matching strategy of Honoré and Kyriazidou (2000) is inapplicable. Of course, other identification opportunities may arise under a weaker set of conditions as the next section illustrates.

4.7 Identification of the pure AR(2) logit model

The preceding sections introduced a methodology in the AR(p) framework with $p > 1$ to derive moment functions that are valid for all possible realizations of the regressors and initial condition. Here, we show how we can leverage these moment functions to formally prove identification of the common parameters of purely autoregressive models, focusing for simplicity of exposition on the AR(2).

For the case $T = 3$, Honoré and Weidner (2020) showed with the assistance of a computer algebra system (e.g Mathematica) that γ_1 and γ_2 are identified if some combinations of initial conditions (Y_{i0}, Y_{i-1}) are observed with strictly positive probability. We use our simple two-step approach to re-derive this result purely analytically and offer new insights. With $T = 3$, we know that the transition functions associated to: $\pi^{0|0,0}(A_i), \pi^{0|1,1}(A_i), \pi^{1|1,0}(A_i), \pi^{1|1,1}(A_i)$ can be computed with formulas provided by Theorem 1. After some algebraic simplifications, they read:

$$\phi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} - (1 - e^{\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}})Y_{i2}e^{\gamma_2 Y_{i3} Y_{i0}}(1 - Y_{i1})$$
$$\phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} + (1 - e^{\gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})(1 - Y_{i2})e^{Y_{i3}(\gamma_1 - \gamma_2(1 - Y_{i0}))}Y_{i1}$$
$$\phi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = Y_{i1} + (1 - e^{-(\gamma_1 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}))})Y_{i2}e^{(1 - Y_{i3})(\gamma_1 - \gamma_2 Y_{i0})}(1 - Y_{i1})$$
$$\phi_\theta^{1|1,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - (1 - Y_{i1}) - (1 - e^{\gamma_1 + \gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})\left(1 - Y_{i2}e^{\gamma_2(1 - Y_{i3})(1 - Y_{i0})}\right)Y_{i1}$$

Because $T < 4$, the arguments developed previously are not directly applicable. Yet, the logic to construct identifying moments remains the same: we want to find another set of transition functions to formulate a differencing strategy. With fewer observations then $T = 2 + p$ (here $p = 2$), we will show how this can be achieved by leveraging subsets of the four possible initial conditions in the population: $\{(0,0),(0,1),(1,0),(1,1)\}$.

The solution to complete **Step 1)** and then **Step 2)** comes from noticing that by Theorem 1 (more precisely Lemma 11 used in its proof), we also have

$$\phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = (1 - Y_{i1})e^{Y_{i2}\left(\gamma_1 Y_{i0} - \gamma_2(Y_{i0} - Y_{i-1})\right)}$$
$$\phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = Y_{i1}e^{(1 - Y_{i2})\left(\gamma_1(1 - Y_{i0}) + \gamma_2(Y_{i0} - Y_{i-1})\right)}$$
$$E\left[\phi_{\theta_0}^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0)|Y_{i-1}, Y_{i0}, A_i\right] = \pi^{0|0,Y_{i0}}(A_i)$$
$$E\left[\phi_{\theta_0}^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0)|Y_{i-1}, Y_{i0}, A_i\right] = \pi^{1|1,Y_{i0}}(A_i)$$

Therefore, if $Y_{i0} = 0$, $\phi_\theta^{0|0}(Y_{i1}^2, 0, Y_{i-1})$ and $\phi_\theta^{1|1}(Y_{i1}^2, 0, Y_{i-1})$ are also transition functions associated to $\pi^{0|0,0}(A_i)$ respectively $\pi^{1|1,0}(A_i)$. This observation suggests that for the initial condition $Y_{i0} = 0$, two valid moment functions are:

$$\psi_\theta^{0|0,0}(Y_{i1}^3, 0, Y_{i-1}) = \phi_\theta^{0|0,0}(Y_{i1}^3, 0, Y_{i-1}) - \phi_\theta^{0|0}(Y_{i1}^2, 0, Y_{i-1})$$
$$\psi_\theta^{1|1,0}(Y_{i1}^3, 0, Y_{i-1}) = \phi_\theta^{1|1,0}(Y_{i1}^3, 0, Y_{i-1}) - \phi_\theta^{1|1}(Y_{i1}^2, 0, Y_{i-1})$$

And if $Y_{i0} = 1$, $\phi_\theta^{0|0}(Y_{i1}^2, 1, Y_{i-1})$ and $\phi_\theta^{1|1}(Y_{i1}^2, 1, Y_{i-1})$ are then transition functions associated to $\pi^{0|0,1}(A_i)$ and $\pi^{1|1,1}(A_i)$ respectively. This suggests that for the initial condition $Y_{i0} = 1$, two valid moment functions are:

$$\psi_\theta^{0|0,1}(Y_{i1}^3, 1, Y_{i-1}) = \phi_\theta^{0|0,1}(Y_{i1}^3, 1, Y_{i-1}) - \phi_\theta^{0|0}(Y_{i1}^2, 1, Y_{i-1})$$
$$\psi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1}) = \phi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1}) - \phi_\theta^{1|1}(Y_{i1}^2, 1, Y_{i-1})$$

Straightforward calculations detailed in Appendix Section H show that $\psi_\theta^{0|0,0}(Y_{i1}^3, 0, Y_{i-1})$ and $\psi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1})$ are identically zero and hence are uninformative about γ_{01}, γ_{02}. After a suitable rescaling (see again Appendix Section H), the other two moment functions now denoted with a "tilde" superscript to reflect the normalization read:

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i1}^3, 1, Y_{i-1}) = e^{\gamma_2(1-Y_{i-1})}Y_{i1}(1-Y_{i2})Y_{i3} + e^{-\gamma_1+\gamma_2(1-Y_{i-1})}Y_{i1}(1-Y_{i2})(1-Y_{i3}) - (1-Y_{i1})Y_{i2}$$
$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, Y_{i-1}) = e^{\gamma_2 Y_{i-1}}(1-Y_{i1})Y_{i2}(1-Y_{i3}) + e^{-\gamma_1+\gamma_2 Y_{i-1}}(1-Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1-Y_{i2})$$

For the initial condition $Y_{i0} = 1, Y_{i-1} = 1$ respectively $Y_{i0} = 0, Y_{i-1} = 0$, $\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, 1, 1)$ respectively $\widetilde{\psi_\theta^{1|1,0}}(Y_{i3}, Y_{i2}, 0, 0)$ only depend on γ_1 and are strictly monotonic in this parameter. Therefore, $E\left[\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, 1, 1)|Y_{i-1} = 1, Y_{i0} = 1\right]$ and $E\left[\widetilde{\psi_\theta^{1|1,0}}(Y_{i3}, Y_{i2}, 0, 0)|Y_{i-1} = 0, Y_{i0} = 0\right]$ point identify γ_{01}. Likewise, for the initial condition $Y_{i0} = 1, Y_{i-1} = 0$ respectively $Y_{i0} = 0, Y_{i-1} = 1$, $\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, 1, 0)$ respectively $\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, Y_{i-1})$ are strictly monotonic in γ_2. It follows that provided that γ_{01} is identified, $E\left[\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, 1, 0)|Y_{i-1} = 0, Y_{i0} = 1\right]$ and $E\left[\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, 1)|Y_{i-1} = 1, Y_{i0} = 0\right]$ point identify γ_{02}. Taking stock, if $P(Y_{i-1} = 1, Y_{i0} = 1) > 0$ or $P(Y_{i-1} = 0, Y_{i0} = 0) > 0$, γ_{01} is identified and if we also have $P(Y_{i-1} = 0, Y_{i0} = 1) > 0$ or $P(Y_{i-1} = 1, Y_{i0} = 0) > 0$, γ_{02} is also identified (see also Honoré and De Paula (2021) for a similar discussion).

However, in the absence of sufficient variation in the initial condition, it is clear that this line of arguments fails to identify all the common parameters. In fact, for a fixed initial condition in the population, Dobronyi et al. (2021) formally prove that the common parameters are only set identified. In that case, $T = 4$ is actually the minimum number of periods required to identify $\theta_0 = (\gamma_{01}, \gamma_{02})$. This follows for example from the fact that with $T = 4$, γ_{02} is identified (Chamberlain (1985)) and the fact that for any or the four initial conditions, the moment functions presented above are always strictly monotonic in γ_1 and hence identify the true parameter given that γ_{02} is identified.

4.8 Identification of the AR(2) logit model with strictly exogenous regressors

Next, we discuss point identification in the AR(2) logit model with strictly exogenous explanatory variables.

As in the pure version of the model, for $T = 3$, the identification argument relies on matching the transition probabilities across period $t = 1$ and $t = 2$ in order to construct valid moment functions. The only difference is that since the transition probabilities are covariate-dependent, the approach further requires $X_{i2} = X_{i3}$. Then, a direct generalization of the steps laid out in the pure model yield two valid moment functions for the initial conditions $Y_{i0} = 1$ and $Y_{i0} = 0$ given by

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i1}^3, 1, Y_{i-1}) = e^{\gamma_2(1-Y_{i-1})+X'_{i31}\beta} Y_{i1}(1-Y_{i2})Y_{i3} + e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X'_{i31}\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3}) - (1-Y_{i1})Y_{i2}$$

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, Y_{i-1}) = e^{\gamma_2 Y_{i-1}-X'_{i31}\beta}(1-Y_{i1})Y_{i2}(1-Y_{i3}) + e^{-\gamma_1+\gamma_2 Y_{i-1}-X'_{i31}\beta}(1-Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1-Y_{i2})$$

In addition to being strictly monotonic in γ_1 and γ_2, Honoré and Weidner (2020) observe that they are strictly increasing or decreasing in β_k depending on whether $X_{ik,3} > X_{ik,1}$ or $X_{ik,3} < X_{ik,1}$. They show that this particular feature can be fruitfully exploited if the initial conditions vary in the population to prove that the model parameters are identified (see their Theorem 2).

However, it is important to stress that as in Honoré and Kyriazidou (2000), the requirement that we be able to match each explanatory variables in $t = 2$ and $t = 3$ is the key component of this argument and has some issues. One issue relates to the "curse of dimensionality": the rate of convergence of kernel-type estimators drawing on these moments deteriorates with the number of continous components of X_{it}. The second issue, mentioned previously, is that it precludes time-specific effects. In fact, in the presence of such regressors, the analysis of Dobronyi et al. (2021) revealed that there are actually no moment equality conditions available in the model. Thus, to our understanding, the two questions of i) parameter identifiability in models with more than a single lag and free-varying covariates and ii) whether we can improve upon existing results remain unclear.

A logic next step is to consider the same model with an additional period of observation, i.e $T = 4$, where the advantage is that the valid moment functions of Proposition 3 - suiting any type of strictly exogenous regressors - become available. Their full expressions, reported in Appendix Section I.2 are nevertheless particularly intricate and as such not immediately helpful to make progress on the identification of $\theta_0 = (\gamma_{01}, \gamma_{02}, \beta'_0)$. In light of this, one default option would be to compute the identified set implied by our moment equality conditions or by the model likelihood following Dobronyi et al. (2021). Another option pursued here is to explore the usefulness of additional identifying assumptions that may be of plausible empirical relevance. We consider most specifically the case where a 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 $Z_i = (R'_i, W_{i1}, W_{i3}, W_{i4})$, $X_{it} = (W_{it}, R'_{it}) \in \mathbb{R}^{K_x}$ 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 accomodate time effects in the baseline one lag model.

Assumption 1. *(i) The covariate W_{i2} has 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 contexts, e.g the econometrician may have a confident prior that say income affects tobacco/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/or 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) $\lim_{w_2 \to \infty} p(a|y^0, z, w_2) = q(a|y^0, z), \ \lim_{w_2 \to \infty} p(z|y^0, w_2) = q(z|y^0)$

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

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

Assumption 2 provides regularity conditions 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 3 reduces to

$$\begin{aligned}\psi_{\theta,\infty}^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, Z_i) = &-(1 - Y_{i1})(1 - Y_{i2})Y_{i3} \\ &+ \left[e^{X_{i34}'\beta} - 1\right](1 - Y_{i1})(1 - Y_{i2})(1 - Y_{i3})Y_{i4} \\ &+ e^{-\gamma_1 Y_{i0} + \gamma_2(1 - Y_{i-1}) + X_{i31}'\beta} Y_{i1}(1 - Y_{i2})(1 - Y_{i3})Y_{i4} \\ &+ e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X_{i41}'\beta} Y_{i1}(1 - Y_{i2})(1 - Y_{i3})(1 - Y_{i4})\end{aligned} \tag{10}$$

which happens 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. To this end, let $(\bar{x}, \underline{x}) \in \mathbb{R}^2$, such that $\bar{x} > \underline{x}$ and define the sets

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

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 3. *For $T = 4$, suppose that outcomes $(Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ are generated from model (9) with $p = 2$, initial condition $y^0 \in \{0,1\}^2$, common parameters $\theta_0 = (\gamma_{10}, \gamma_{20}, \beta_0') \in \mathbb{R}^{2+K_x}$ 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 $y^0 \in \{0,1\}^2$*

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

Let

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

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

$$\Psi_{s,y^0}^{0|0,0}(\theta) = 0, \quad \forall s \in \{-,+\}^{K_x}, \quad \forall y^0 \in \{0,1\}^2$$

Theorem 3 is compelling because we formally learn that point identification 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. 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 $T \geq 2 + p$.

On the more practical front, we note as in Dobronyi et al. (2023) that an estimator of θ_0 based on Theorem 3 will be immune to the curse of dimensionality inherent to matching strategies. The implied rate of convergence will be independent of the number of regressors albeit slower than $\sqrt{N}$ as is typical of estimators exploiting *irregular identification*. For this reason, we actually discourage using the moments of Theorem 3 for estimation purposes even in a context where Assumption 1 may be reasonable. Our practical recommendation is to proceed as we do in the simulations of Section 6 and use the valid moment functions of Proposition 3. Despite their complex form, they should "generally" lead to consistent estimators converging at the parametric rate.

4.9 Average Marginal Effects in the AR(2)

Even though the distribution of unobserved heterogeneity cannot be identified in short panels, we have previously shown that AMEs of the AR(1) model are in fact identified if four consecutive periods are available. This section generalizes this insight to the AR(2) model and we illustrate the usefulness of this extension to measure more refined causal quantities.

With two lags, there are eight total average transition probabilities for any subpopulation of individuals with covariates x in any given period t. They read,

$$\Pi_t^{k|l_1,l_2}(y_{-1}, y_0, x) = E\left[\pi_t^{k|l_1,l_2}(X_i, A_i)|Y_{i-1} = y_{-1}, Y_{i0} = y_0, X_i = x\right] = \int \pi_t^{k|l_1,l_2}(x,a)p(a|y_{-1}, y_0, x)da$$

for all $(k, l_1, l_2) \in \mathcal{Y}^3$, and $x \in \mathbb{R}^{TK}$. It follows that a richer set of AMEs becomes available compared to the baseline. To fix ideas, suppose that $Y_{it} \in \{0, 1\}$ is a dummy for employment status in period t. Then, the econometrician may be interested in the causal effect on future employment of transitioning out of unemployment in the past for both currently employed and unemployed individuals; two quantities we cannot adequately capture with a one lag specification. With the AR(2), this would correspond to:

$$\Pi_t^{1|1,1}(y_{-1}, y_0, x) - \Pi_t^{1|1,0}(y_{-1}, y_0, x) \text{ and } \Pi_t^{1|0,1}(y_{-1}, y_0, x) - \Pi_t^{1|0,0}(y_{-1}, y_0, x)$$

The significance of these two AMEs will inform the econometrician on the importance of past experiences for labor market success. The relative magnitude of γ_{01} and γ_{02} will determine the sign of these estimands but will not pin down their magnitude which partially depends on the distribution of latent heterogeneity. This is what makes the separate identification analysis of AMEs particularly relevant (Aguirregabiria and Carro (2021)). Of course, other types of contrasts may be considered and more flexible specifications with 3 lags and beyond may also be useful to adress certain questions.

In the context of the AR(2) model, we know from the previous section that the structural parameters, namely $\theta_0 = (\gamma_{01}, \gamma_{02}, \beta_0')'$ will be identified with discrete covariates as long as $T \geq 4$ ($T = 3$ if there is variation in the initial condition). This implies that the average transition functions are identified, being expectations of the model transition functions. Specifically, with discrete regressors if $T \geq 4$, for all $t \in \{2, \ldots, T-1\}$ and for all $(y_1, y_2) \in \{0, 1\}^2$

$$\Pi_t^{y_1|y_1,y_2}(y_{-1}, y_0, x) = E\left[\phi_{\theta_0}^{y_1|y_1,y_2}(Y_{it+1}, Y_{it}, Y_{it-3}^{t-1}, X_i)|Y_{i-1} = y_{i-1}, Y_{i0} = y_0, X_i = x\right]$$

$$\Pi_t^{1-y_1|y_1,y_2}(y_{-1}, y_0, x) = 1 - \Pi_t^{y_1|y_1,y_2}(y_{-1}, y_0, x)$$

One may be also compute these quantities with the transition functions of Lemma 4 and Corollary 4.1 in long enough panels. In turn, AMEs which are contrasts of average transition functions are point identified.

5 Multi-dimensional fixed effects models

We now turn our attention to multi-dimensional fixed effects models for which fewer methods of estimation are currently discussed in the literature in the presence of strictly exogenous covariates. We show that the general blueprint developed in the scalar case to derive valid moment functions carries over to VAR(1) and MAR(1) models.

5.1 Moment conditions for the VAR(1) logit with exogenous regressors

We begin with the analysis of VAR(1) logit models, variants of which have been successfully used to study the relationship between sickness and unemployment (Narendranthan et al. (1985)), the progression from softer drug use to harder drug use among teenagers (Deza (2015)), transitivity in networks (Graham (2013)) and more recently the employment of couples (Honoré et al. (2022)). For a given $M \geq 2$, the model reads:

$$Y_{m,it} = \mathbb{1}\left\{\sum_{j=1}^{M} \gamma_{0mj} Y_{j,it-1} + X'_{m,it}\beta_{0m} + A_{m,i} - \epsilon_{m,it} \geq 0\right\}, \quad m = 1, \dots, M, \quad t = 1, \dots, T \tag{11}$$

We let $Y_{it} = (Y_{1,it}, \dots, Y_{M,it})'$ denote the outcome vector in period t with support $\mathcal{Y} = \{0,1\}^M$ of cardinality 2^M. We let $X_{it} = (X'_{1,it}, \dots, X'_{M,it})' \in \mathbb{R}^{K_1} \times \dots \times \mathbb{R}^{K_M}$ denote the vector of exogenous covariates in period t and $A_i = (A_{1,i}, \dots, A_{M,i})' \in \mathbb{R}^M$. The initial condition is now given by $Y_{i0} = (Y_{1,i0}, \dots, Y_{M,i0})' \in \mathcal{Y}$ and the model transition probabilities are given by:

$$\pi_t^{k|l}(A_i, X_i) = P(Y_{it+1} = k | Y_{it} = l, X_i, A_i) = \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{0mj} l_j + X'_{m,it+1}\beta_{0m} + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{0mj} l_j + X'_{m,it+1}\beta_{0m} + A_{m,i}}}$$

for all $(k, l) \in \mathcal{Y} \times \mathcal{Y}$.

Building on Honoré and Kyriazidou (2000), Honoré and Kyriazidou (2019) use a conditional likelihood approach to prove the identification of the structural parameter $\theta_0 = (\gamma_{011}, \gamma_{012}, \gamma_{021}, \gamma_{022}, \beta_{01}, \beta_{02})$ for the bivariate specification when $T = 3$ and the regressors do not vary over the last two periods. As in scalar models, we show hereinbelow that this strong restriction which can yield undesirable rates of convergence is unnecessary to obtain valid moment conditions.

Step 1) in the VAR(1) logit model has a nuance relative to its scalar counterpart in that the only transition functions available are those associated to $\pi_t^{k|k}(A_i, X_i)$, for $k \in \mathcal{Y}$, i.e the probabilities of staying in the same state. As in the AR(1) model, one can determine their expressions by looking for functions of the form $\phi_\theta(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ that: 1) are non-zero only for a single realization of the period t outcome and 2) have a conditional expectation given past outcomes, regressors and the fixed effect equal to one of the 2^M target transition probabilities. This procedure is particularly straightforward to implement in the bivariate VAR(1) as we detail in Appendix Section J. Once all four transition functions are obtained for the case $M = 2$, it becomes clear that the general functional form is as per Lemma 5. It is then a matter of brute force calculation to verify that this is indeed correct.

Lemma 5. *In model (11) with $T \geq 2$ and $t \in \{1, \dots, T-1\}$, let for all $k \in \mathcal{Y}$*

$$\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = \mathbb{1}\{Y_{it} = k\} e^{\sum_{m=1}^M (Y_{m,it+1} - k_m)\left(\sum_{j=1}^M \gamma_{mj}(Y_{j,it-1} - k_j) - \Delta X'_{m,it+1}\beta_m\right)}$$

Then:

$$E\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) | Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{k|k}(A_i, X_i) = \prod_{m=1}^{M} \frac{e^{k_m(\sum_{j=1}^M \gamma_{0mj} k_j + X'_{m,it+1}\beta_{0m} + A_{m,i})}}{1 + e^{\sum_{j=1}^M \gamma_{0mj} k_j + X'_{m,it+1}\beta_{0m} + A_{m,i}}}$$

To obtain identifying moments for the model parameters, we now proceed in a similar vein to the AR(1) model with exogenous regressors. First, we start by noticing that when $T \geq 3$ and for $t \in \{2, \dots, T-1\}$, one can find transition functions $\zeta_\theta^{k|k}(.)$ different from $\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$ also associated to $\pi_t^{k|k}(A_i, X_i)$ for any $k \in \mathcal{Y}$. This property is an immediate consequence of our second partial fraction decomposition formula in Appendix

Lemma 10 which we may regard as a generalization of Kitazawa (2022)'s hyperbolic transformations to the multivariate case. Owing to the special logistic structure of the model probabilities, these additional transition functions $\zeta_\theta^{k|k}(.)$ can be found by considering a linear combination of $\mathbb{1}(Y_{is} = k)$, with $s \in \{1, \dots, t-1\}$ and the sum over all $l \in \mathcal{Y} \setminus \{k\}$ of interactions between $\mathbb{1}(Y_{is} = l)$ and any transition functions having no dependence on outcomes prior to period s, e.g $\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$. Lemma 6 below gives the particular transition functions that can be constructed directly from those of Lemma 5.

Lemma 6. *In model (11) with $T \geq 3$, for all t, s such that $T-1 \geq t > s \geq 1$, let for all $m \in \{1, \dots, M\}$ and $(k, l) \in \mathcal{Y}^2$*

$$\mu_{m,s}(\theta) = \sum_{j=1}^{M} \gamma_{mj} Y_{j,is-1} + X'_{m,is}\beta_m$$

$$\kappa_{m,t}^{k|k}(\theta) = \sum_{j=1}^{M} \gamma_{mj} k_j + X'_{m,it+1}\beta_m$$

$$\omega_{t,s,l}^{k|k}(\theta) = 1 - e^{\sum_{j=1}^{M}(l_j - k_j)\left[\kappa_{j,t}^{k|k}(\theta) - \mu_{j,s}(\theta)\right]}$$

and define the moment functions

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

Then,

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

Beyond $T = 4$, more transition functions are available and can be derived sequentially from those of Lemma 6. This of course is not surprising in light of the results obtained for the AR(1) model. See Corollary 6.1 for their expressions.

Corollary 6.1. *In model (11) with $T \geq 4$, for any t and ordered collection of indices s_1^J, $J \geq 2$, satisfying $T-1 \geq t > s_1 > \dots > s_J \geq 1$, let for all $k \in \mathcal{Y}$*

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

with weights $\omega_{t,s_J,l}^{k|k}(\theta)$ defined as in Lemma 6. Then,

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

One can then obtain a family of valid moment functions for **Step 2)** by adequately repurposing the statement of Proposition 2 to the VAR(1) case, i.e by updating the expressions of $\phi_\theta^{k|k}(.)$ and $\zeta_\theta^{k|k}$ according to Lemma 5 and Corollary 6.1. To conserve on space and avoid repetition, we leave this simple exercise to the reader.

Remark 6. Similarly to Remarks 4 and 5 for the scalar case, we emphasize that the tools developed here can be modified to handle other interesting variants featuring more complex interdependencies across the different layers of the model indexed by $m = 1, \dots, M$. To illustrate the wider applicability of our methodology, we show in Appendix L how one can derive moment restrictions in the dynamic network formation model of Graham (2013) and extensions thereof incorporating exogenous covariates.

5.2 Moment conditions for the MAR(1) logit with exogenous regressors

Last, we cover dynamic multinomial logit models which have been utilized to measure state-dependence in a range of economic contexts including: employment history in the French labor market (Magnac (2000)), the impact of international trade on the transition matrix of employment across sectors (Egger et al. (2003)) and consumer product choice (Dubé et al. (2010)) amongst others. Dubé et al. (2010) actually pursue a random effect approach but our focus here is entirely on a fixed effect setting.

We focus on the the baseline MAR(1) logit model with fixed effects introduced by Magnac (2000) and further extended in Honoré and Kyriazidou (2000). The model assumes a fixed number of alternatives $C+1$ with $C \geq 1$ and is characterized by the following transition probabilities:

$$\pi_t^{k|l}(A_i, X_i) = P(Y_{it+1} = k | Y_{it} = l, X_i, A_i) = \frac{e^{\gamma_{kl} + X'_{ikt+1}\beta_k + A_{ik}}}{\sum_{c=0}^{C} e^{\gamma_{cl} + X'_{ict+1}\beta_j + A_{ic}}}, \quad t = 1, \dots, T \tag{12}$$

with $(k, l) \in \mathcal{Y} = \{0, 1, \dots, C\}$. Here, $Y_{it} \in \mathcal{Y}$ indicates the choice of individual i in period t, X_{ijt} denotes a vector of individual-alternative specific exogenous covariates and $A_{ij} \in \mathbb{R}$ is the fixed effect attached to alternative j for individual i. The initial condition is $Y_{i0} \in \mathcal{Y}$ and in keeping with the fixed effect assumption, its conditional distribution given unobserved heterogeneity and the regressors, $\big(P(Y_{i0} = k | X_i, A_i)\big)_{k=1}^{C}$, is left fully unrestricted. Following Magnac (2000), we normalize the transition parameters and fixed effect of the reference alternative "0" to zero [8]. That is $\gamma_{j0} = \gamma_{0j} = 0, A_{0,j} = 0$ for all $j \in \mathcal{Y}$ leaving $\theta = \big((\gamma_{kl})_{k,l \geq 1}, (\beta_l)_{l \geq 0}\big)$ as the unknown model parameters.

This specification can be motivated by assuming that agents rank options according to random latent utility indices with disturbances independent over time and across alternatives. In this context, equation (12) is obtained if the best alternative is selected and the error terms are Type 1 extreme value distributed conditional on Y_{i0}, A_i, X_i (see the simulation design in Section 6). Magnac (2000) studies the "pure" case without covariates and shows 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) show that this argument carries over to the case with exogenous explanatory variables if one matches the regressors across specific time periods. Here, we offer an alternative estimation strategy that circumvents the need for matching.

Similarly to the VAR(1) model, **Step 1)** for the MAR(1) is only possible for the transition probabilities

[8] The transition 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 $A^*_{ik} = A_{ik} - A_{i0}$.

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 facilitate the derivation of the relevant transition functions, we follow our usual heuristic of looking for $\phi_\theta^{k|k}(.), k \in \mathcal{Y}$ satisfying:

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

Upon obtaining their exact expressions for the simplest case with $C = 2$, it is easy to conjecture and verify by direct calculations that the general expressions of the $C + 1$ transition functions of the MAR(1) model are as displayed in Lemma 7.

Lemma 7. *In model (12) 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:

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

Unsurprisingly, given the similarities shared between the MAR(1) and all other specifications discussed in the paper, so long as $T \geq 3$, one can again derive transition functions other than $\phi_\theta^{k|k}(Y_{it-1}^{t+1}, X_i)$ also associated to $\pi_t^{k|k}(A_i, X_i)$ for $k \in \mathcal{Y}$ in periods $t \in \{1, \ldots, T-1\}$. The simple logistic identities of Appendix Lemma 9 imply that these transition functions, that we keep denoting $\zeta_\theta^{k|k}(.)$ have a similar form to those of the VAR(1) model as shown in Lemma 8.

Lemma 8. *In model (12) 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

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

Then,

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

Additionally, if the econometrician has access to a dataset with more than four observations per sampling unit - counting the initial condition - then, more transition functions associated to the same transition probabilities are available per Corollary 8.1.

Corollary 8.1. *In model (12) with $T \geq 4$, for any t and ordered collection of indices s_1^J, $J \geq 2$, satisfying $T-1 \geq t > s_1 > \ldots > s_J \geq 1$, let for all $k \in \mathcal{Y}$*

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

with weigts $\omega_{t,s_J,l}^{k|k}(\theta)$ defined as in Lemma 8. Then,

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

This completes **Step 1)** for the MAR(1) logit model. For **Step 2)**, we recommend a family of valid moment functions mirroring those of Proposition 2 for the AR(1) case to ensure their linear independence. Note that the same differencing strategy works - with the MAR(1) transition functions - by simple virtue of the law of iterated expectations.

6 Simulation Experiments

In this section, we report the results of a small set of simulations designed to assess the finite sample performance of GMM estimators based on our moment conditions.

6.1 Monte Carlo for an AR(3) logit model

For our first example, we consider an AR(3) logit model with $T = 5$ periods (i.e 8 periods in total with the initial condition) and a single exogenous covariate. We set the common parameters to $\gamma_{01} = 1.0$, $\gamma_{02} = 0.5, \gamma_{03} = 0.25, \beta_0 = 0.5$ and use the following generative model in the spirit of Honoré and Kyriazidou (2000):

$$\begin{aligned}
Y_{i-2} &= \mathbb{1}\{X_{i-2}'\beta_0 + A_i - \epsilon_{i-2} \geq 0\} \\
Y_{i-1} &= \mathbb{1}\{\gamma_{01}Y_{i-2} + X_{i-1}'\beta_0 + A_i - \epsilon_{i-1} \geq 0\} \\
Y_{i0} &= \mathbb{1}\{\gamma_{01}Y_{i-1} + \gamma_{02}Y_{i-2} + X_{i0}'\beta_0 + A_i - \epsilon_{i0} \geq 0\} \\
Y_{it} &= \mathbb{1}\left\{\gamma_{01}Y_{it-1} + \gamma_{02}Y_{it-2} + \gamma_{03}Y_{it-3} + X_{it}'\beta_0 + A_i - \epsilon_{it} \geq 0\right\}, \quad t = 1, \ldots, 5
\end{aligned}$$

The disturbances ϵ_{it} are iid standard logistic over time, X_{it} is iid $\mathcal{N}(0,1)$ and the fixed effects are computed as $A_i = \frac{1}{\sqrt{8}} \sum_{t=-2}^{5} X_{it}$. To evaluate the performance of the estimators described below, we simulate data for four sample sizes : 500, 2000, 8000, 16000, and perform 1000 Monte Carlo replications for each design.

For $T = 5$, we know from Proposition 3 that 8 valid moment functions are available, each stemming from the 8 possible transition probabilities of the model (there are really 16 transition probabilities in total but 8 are redundant since probabilities sum to one). Their full expressions are provided in Appendix Section O.1.We consider the interaction of all 8 valid moment functions with the 3 initial conditions Y_{i-2}, Y_{i-1}, Y_{i0}

and the difference in covariates $X_{i5} - X_{it}$ for $t \in \{1, \ldots, 4\}$ to construct the 56×1 moment vector:

$$m_\theta(Y_i, Y_i^0, X_i) = \begin{pmatrix} \psi_\theta^{0|0,0,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{0|0,0,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{0|0,1,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{0|0,1,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{1|1,0,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{1|1,0,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{1|1,1,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) \\ \psi_\theta^{1|1,1,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) \end{pmatrix} \otimes \begin{pmatrix} Y_{i-2} \\ Y_{i-1} \\ Y_{i0} \\ X_{i5} - X_{i1} \\ X_{i5} - X_{i2} \\ X_{i5} - X_{i3} \\ X_{i5} - X_{i4} \end{pmatrix}$$

where $\otimes$ denotes the standard Kronecker product. The choice of this particular set of intruments is motivated both by simplicity and by the fact that the model's valid moment functions depends on X_i only through the differences $X_{i5} - X_{it}$ for periods $t \in \{1, \ldots, 4\}$. We also consider a rescaled version of $m_\theta(Y_i, Y_i^0, X_i)$ that we denote $\widetilde{m_\theta}(Y_i, Y_i^0, X_i)$ where each of the 8 valid moment functions are appropriately rescaled so that $\forall y_1^3 \in \{0,1\}^3$, $\sup_{X_i, Y_i, \theta} \left|\psi_\theta^{y_1|y_1,y_2,y_3}(Y_{i-1}^5, Y_{i-2}^1, X_i)\right| < \infty$. We do so by normalizing $\psi_\theta^{y_1|y_1,y_2,y_3}(Y_{i-1}^5, Y_{i-2}^1, X_i)$ by the sum of the absolute values of all unique values it can take as a function over choice histories Y_{i1}^5. The rationale for normalizing the moments originates from Honoré and Weidner (2020) who presented numerical evidence that a rescaling of this kind improved the finite sample performance of their estimators in the one and two lags cases. Given, $m_\theta(Y_i, Y_i^0, X_i)$ and $\widetilde{m_\theta}(Y_i, Y_i^0, X_i)$, we study the properties of 3 different GMM estimators:

$$\hat{\theta}^a = \arg\max_{\theta \in \mathbb{R}^4} \left(\frac{1}{N}\sum_{i=1}^N m_\theta(Y_i, Y_i^0, X_i)\right)' \left(\frac{1}{N}\sum_{i=1}^N m_\theta(Y_i, Y_i^0, X_i)\right)$$

$$\hat{\theta}^b = \arg\max_{\theta \in \mathbb{R}^4} \left(\frac{1}{N}\sum_{i=1}^N \widetilde{m_\theta}(Y_i, Y_i^0, X_i)\right)' \left(\frac{1}{N}\sum_{i=1}^N \widetilde{m_\theta}(Y_i, Y_i^0, X_i)\right)$$

$$\hat{\theta}^c = \arg\max_{\theta \in \mathbb{R}^4} \left(\frac{1}{N}\sum_{i=1}^N \widetilde{m_\theta}(Y_i, Y_i^0, X_i)\right)' W(\hat{\theta}^b) \left(\frac{1}{N}\sum_{i=1}^N \widetilde{m_\theta}(Y_i, Y_i^0, X_i)\right)$$

where $W(\hat{\theta}^b)$ is a diagonal matrix with its diagonal entries equal to the inverse of the sample variances of each entry of $\widetilde{m_{\hat{\theta}^b}}(Y_i, Y_i^0, X_i)$. Note that $\hat{\theta}^c$ is then a two-step estimator since $\hat{W}(\hat{\theta}^b)$ depends upon $\hat{\theta}^b$. Moreover, while the first two GMM estimators assign equal weight to all moments, the third downweights the least informative in an attempt to gain in precision. Under standard regularity conditions, $\hat{\theta}^a, \hat{\theta}^b, \hat{\theta}^c$ should be consistent and asymptotically normal.

Table 1 presents the median bias and median absolute errors of the three GMM estimators for each design $N \in \{500, 2000, 8000, 16000\}$. Figure 2 in Appendix Section O.2 plots their densities which as expected resemble gaussian distributions for the larger values of N. Interestingly, a first observation is that all estimators appear to suffer from a negative bias at least up to $N = 8000$. And while this bias effectively vanishes for the

Table 1: Performance of GMM estimators for the AR(3)

		$\hat{\gamma_1}^a$	$\hat{\gamma_1}^b$	$\hat{\gamma_1}^c$	$\hat{\gamma_2}^a$	$\hat{\gamma_2}^b$	$\hat{\gamma_2}^c$	$\hat{\gamma_3}^a$	$\hat{\gamma_3}^b$	$\hat{\gamma_3}^c$	$\hat{\beta}^a$	$\hat{\beta}^b$	$\hat{\beta}^c$
$N = 500$													
	Bias	-0.56	-0.81	-0.81	-0.51	-0.33	-0.42	-0.37	-0.45	-0.33	-0.18	-0.05	-0.03
	MAE	0.56	0.92	0.91	0.51	0.54	0.53	0.37	0.49	0.37	0.18	0.10	0.10
$N = 2000$													
	Bias	-0.45	-0.42	-0.33	-0.46	-0.15	-0.16	-0.30	-0.21	-0.09	-0.10	-0.01	-0.00
	MAE	0.45	0.60	0.57	0.46	0.34	0.31	0.30	0.29	0.18	0.10	0.05	0.05
$N = 8000$													
	Bias	-0.30	-0.06	-0.02	-0.35	-0.04	-0.02	-0.22	-0.06	-0.03	-0.06	-0.01	-0.00
	MAE	0.30	0.25	0.23	0.35	0.15	0.13	0.22	0.12	0.08	0.06	0.03	0.03
$N = 16000$													
	Bias	-0.24	-0.01	0.01	-0.29	-0.01	-0.00	-0.18	-0.02	-0.01	-0.04	-0.00	-0.00
	MAE	0.25	0.17	0.15	0.29	0.11	0.10	0.18	0.08	0.05	0.04	0.02	0.02

NOTES: *Bias and MAE stand for median bias and median absolute error respectively. Reported results are based on a 1000 replications of the DGP.*

"rescaled" GMM estimators for the larger sample size $N = 16000$, it remains quite significant for all lag parameters and the slope coefficient for the "unnormalized" estimator. This is evident from the sign of the bias in Table 1 and from the fact that all green densities are to the left of the true parameters in Figure 2. This observation confirms the practical importance of normalizing all valid moment functions in dynamic fixed effects logit models to obtain precise estimates in small samples. Focusing now on the "rescaled" estimators, we can see that they perform relatively well for $N \geq 8000$ with a small advantage in accuracy for $\hat{\theta}^c$ reflected by the overall smaller median absolute bias. This is corroborated in Figure 2: the blue and red densities are approximately centered at the true parameter values for $N \geq 8000$ and the red density is slightly more concentrated than the blue. Estimates for the slope parameter β are quite accurate even for $N = 500$ but precise estimation of the transition parameters requires a larger sample size. In terms of median absolute bias, it is interesting to note a strict ranking on the precision of estimates of the transition parameters: the coefficient on the first lag is noisier than the coefficient on the second lag which itself is noisier than the coefficient on the third lag for each $N \in \{500, 2000, 8000, 16000\}$. In an unreported set of simulations, we have found that this empirical pattern is robust to other choices of the population parameters and initial condition and also applies to the AR(2) model with a similar data generating process.

6.2 Monte Carlo for a MAR(1) logit model

In our next example, we examine a dynamic multinomial logit model with $M = 2$, $T = 3$ and a scalar regressor X_{ict}, $c \in \{0, 1, 2\}$. We set the common parameters to $\gamma_{011} = \gamma_{022} = 1.0$, $\gamma_{012} = \gamma_{021} = 0.5$, $\beta_{00} = \beta_{01} = \beta_{02} = 0.5$ and impose the normalization $\gamma_{000} = \gamma_{010} = \gamma_{020} = \gamma_{001} = \gamma_{002} = 0$. The data generating process is:

$$
\begin{aligned}
Y^*_{ic0} &= X'_{ic0}\beta_c + A_{ic} + \epsilon_{ic0}, \quad \forall c \in \{0,1,2\} \\
Y_{i0} &= c \text{ if } Y^*_{ic0} = \max_{k=0,1,2} Y^*_{ik0} \\
Y^*_{ict} &= \gamma_{c1}\mathbb{1}(Y_{it-1} = 1) + \gamma_{c2}\mathbb{1}(Y_{it-1} = 1) + X'_{ict}\beta_c + A_{ic} + \epsilon_{ict}, \quad \forall c \in \{0,1,2\}, \quad t = 1,2,3 \\
Y_{it} &= c \text{ if } Y^*_{ict} = \max_{k=0,1,2} Y^*_{ikt}, \quad t = 1,2,3
\end{aligned}
$$

where the disturbances ϵ_{ict} are iid Gumbel(0,1), the covariates X_{ict} are iid $\mathcal{N}(0,1)$ and the fixed effects are computed as $A_{ic} = \frac{1}{\sqrt{4}} \sum_{t=0}^{3} X_{it}$. The specifics of the simulation design are otherwise unchanged.

Since $T = 3$, Section 5.2 in conjunction with Proposition 2 predict three valid moment functions, viz. $\psi_\theta^{0|0}(Y^3_{i1}, Y^1_{i0}, X_i), \psi_\theta^{1|1}(Y^3_{i1}, Y^1_{i0}, X_i), \psi_\theta^{2|2}(Y^3_{i1}, Y^1_{i0}, X_i)$ whose exact expressions can be found in Appendix Section O.3. Given the importance of rescaling these valid moment functions to obtain good finite sample properties for GMM estimators in the context of the AR(3), we normalize each $\psi_\theta^{k|k}(Y^3_{i1}, Y^1_{i0}, X_i)$ for $k \in \{0,1,2\}$ by the sum of the absolute values of their unique non-zero entries as an eight-dimensional vector (8 possible choice histories Y^3_{i1} per initial condition). We denote the resulting moment functions as $\widetilde{\psi_\theta^{k|k}}(Y^3_{i1}, Y^1_{i0}, X_i)$ and interact them with the initial condition Y_{i0} and the differences of covariates: $X_{ict} - X_{ics}$ for all $t > s$ and $c \in \{0,1,2\}$ to produce the 30×1 moment vector

$$
\widetilde{m_\theta}(Y_i, Y^0_i, X_i) = \begin{pmatrix} \widetilde{\psi_\theta^{0|0}}(Y^3_{i1}, Y^1_{i0}, X_i) \\ \widetilde{\psi_\theta^{1|1}}(Y^3_{i1}, Y^1_{i0}, X_i) \\ \widetilde{\psi_\theta^{2|2}}(Y^3_{i1}, Y^1_{i0}, X_i) \end{pmatrix} \otimes \begin{pmatrix} Y_{i0} \\ X_{i03} - X_{i02} \\ X_{i03} - X_{i01} \\ X_{i02} - X_{i01} \\ X_{i13} - X_{i12} \\ X_{i13} - X_{i11} \\ X_{i12} - X_{i11} \\ X_{i23} - X_{i22} \\ X_{i23} - X_{i21} \\ X_{i22} - X_{i21} \end{pmatrix}
$$

Our choice of instruments is guided again by simplicity and the fact that X_i features in each valid moment function only in the form of the above contrasts over time of alternative-specific covariates. With $\widetilde{m_\theta}(Y_i, Y^0_i, X_i)$ in hand, we analyze the finite sample behavior of the MAR(1) analogs of $\hat{\theta}^b$ and $\hat{\theta}^c$ defined previously for the AR(3). The results of the simulations are summarized in Table 2 and Table 3.

As in the AR(3) example, both GMM estimators are rather imprecise in small samples for the transition parameters. This is clear from the magnitude of the median absolute error and to a lesser extent the bias in Table 2 for $N \in \{500, 2000\}$. Additionally, the slope parameters, $\beta_0, \beta_1, \beta_2$ are precisely estimated starting at $N = 2000$. This time however, the general comparison between the unweighted GMM estimator $\hat{\theta}^b$ and the weighted GMM estimator $\hat{\theta}^c$ is less clear: $\hat{\theta}^c$ is significantly more biased for $N \in \{500, 2000\}$ and $\hat{\theta}^b$ seems

Table 2: Performance of GMM estimators for the MAR(1): transition parameters

		$\hat{\gamma_{11}}^b$	$\hat{\gamma_{11}}^c$	$\hat{\gamma_{12}}^b$	$\hat{\gamma_{12}}^c$	$\hat{\gamma_{22}}^b$	$\hat{\gamma_{22}}^c$	$\hat{\gamma_{21}}^b$	$\hat{\gamma_{21}}^c$
$N = 500$									
	Bias	0.02	-0.31	0.01	-0.96	-0.03	-0.38	-0.01	-0.43
	MAE	0.72	1.09	0.70	1.66	0.74	1.27	0.70	1.15
$N = 2000$									
	Bias	-0.10	-0.17	-0.04	-0.27	-0.09	-0.14	-0.06	-0.14
	MAE	0.49	0.65	0.47	0.87	0.47	0.68	0.49	0.64
$N = 8000$									
	Bias	-0.08	-0.08	-0.09	-0.10	-0.11	-0.08	-0.07	-0.07
	MAE	0.29	0.34	0.28	0.34	0.31	0.31	0.29	0.33
$N = 16000$									
	Bias	-0.08	-0.05	-0.07	-0.03	-0.04	-0.01	-0.04	-0.01
	MAE	0.23	0.26	0.20	0.22	0.22	0.22	0.22	0.24

NOTES: *Bias and MAE stand for median bias and median absolute error respectively. Reported results are based on a 1000 replications of the DGP.*

Table 3: Performance of GMM estimators for the MAR(1): slope parameters

		$\hat{\beta_0}^b$	$\hat{\beta_0}^c$	$\hat{\beta_1}^b$	$\hat{\beta_1}^c$	$\hat{\beta_2}^b$	$\hat{\beta_2}^c$
$N = 500$							
	Bias	0.03	0.10	0.02	0.07	0.03	0.09
	MAE	0.12	0.16	0.10	0.14	0.10	0.15
$N = 2000$							
	Bias	0.00	0.02	0.01	0.02	0.01	0.02
	MAE	0.06	0.06	0.05	0.05	0.05	0.05
$N = 8000$							
	Bias	0.01	0.01	0.01	0.01	0.00	0.01
	MAE	0.03	0.03	0.02	0.02	0.02	0.02
$N = 16000$							
	Bias	0.00	0.01	0.00	0.01	0.00	0.00
	MAE	0.02	0.02	0.02	0.02	0.02	0.02

NOTES: *Bias and MAE stand for median bias and median absolute error respectively. Reported results are based on a 1000 replications of the DGP.*

to outperform $\hat{\theta}^c$ in terms of median absolute error across all designs. Also, surprisingly, the bias of $\hat{\theta}^b$ is not monotonic in N as shown in Table 2: it is lower at $N = 500$ then at $N = 8000$ for all transition parameters. We have not investigated these peculiarities - which could be design specific - further at this moment but it is certain that a more thorough analysis of the behavior of GMM in such models would be beneficial. Overall, together with the results for the AR(3), we find that both estimators perform reasonably well in samples of moderate to large size; in accordance with the findings of Honoré and Weidner (2020).

7 Conclusion

We have introduced a systematic algebraic method to construct valid moment functions in dynamic logit models with additive fixed effects and strictly exogenous regressors. It boils down to taking differences between the *transition functions* produced by the model in panels that are sufficiently long. These are distinct functions of the data and the common parameters whose conditional expectation given the initial condition, the covariates and the fixed effect yield the same transition probability. The usefulness of the approach was demonstrated through several examples including AR(p) models where a characterization of the moment conditions is presented for arbitrary $p \geq 1$.

References

Aguirregabiria, V. and Carro, J. M. (2021). Identification of average marginal effects in fixed effects dynamic discrete choice models. *arXiv preprint arXiv:2107.06141*.

Arellano, M. and Honoré, B. (2001). Panel data models: some recent developments. In *Handbook of econometrics*, volume 5, pages 3229–3296. Elsevier.

Bonhomme, S. (2012). Functional differencing. *Econometrica*, 80(4):1337 – 1385.

Browning, M. and Carro, J. M. (2014). Dynamic binary outcome models with maximal heterogeneity. *Journal of Econometrics*, 178(2):805–823.

Card, D. and Hyslop, D. R. (2005). Estimating the effects of a time-limited earnings subsidy for welfare-leavers. *Econometrica*, 73(6):1723–1770.

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

Chay, K. Y., Hoynes, H. W., and Hyslop, D. (1999). A non-experimental analysis of true state dependence in monthly welfare participation sequences. In *American Statistical Association*, pages 9–17.

Chernozhukov, V., Fernández-Val, I., Hahn, J., and Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2):535–580.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.

Davezies, L., D'Haultfoeuille, X., and Laage, L. (2021). Identification and estimation of average marginal effects in fixed effects logit models. *arXiv preprint arXiv:2105.00879*.

Deza, M. (2015). Is there a stepping stone effect in drug use? separating state dependence from unobserved heterogeneity within and between illicit drugs. *Journal of Econometrics*, 184(1):193–207.

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*.

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

Egger, P. H., Pfaffermayr, M., and Weber, A. (2003). Sectoral adjustment of employment: the impact of outsourcing and trade at the micro level. *Available at SSRN 469841*.

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.

Gu, J., Hahn, J., and Kim, K. I. (2023). The information bound of a dynamic panel logit model with fixed effects—corrigendum. *Econometric Theory*, 39(1):219–219.

Hahn, J. (2001). The information bound of a dynamic panel logit model with fixed effects. *Econometric Theory*, 17(5):913–932.

Heckman, J. J. (1981). Heterogeneity and state dependence. In *Studies in labor markets*, pages 91–140. University of Chicago Press.

Honoré, B. E. and De Paula, Á. (2021). Identification in simple binary outcome panel data models. *The Econometrics Journal*, 24(2):C78–C93.

Honoré, B. E., Hu, L., Kyriazidou, E., and Weidner, M. (2022). Simultaneity in binary outcome models with an application to employment for couples. *arXiv preprint arXiv:2207.07343*.

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 Kyriazidou, E. (2019). Panel vector autoregressions with binary data. In *Panel Data Econometrics*, pages 197–223. Elsevier.

Honoré, B. E., Muris, C., and Weidner, M. (2021). Dynamic ordered panel logit models. *arXiv preprint arXiv:2107.03253*.

Honoré, B. E. and Tamer, E. (2006). Bounds on parameters in panel dynamic discrete choice models. *Econometrica*, 74(3):611–629.

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

Kitazawa, Y. (2022). Transformations and moment conditions for dynamic fixed effects logit models. *Journal of Econometrics*, 229(2):350 − 362.

Kitazawa, Y. et al. (2013). Exploration of dynamic fixed effects logit models from a traditional angle. Technical report.

Kitazawa, Y. et al. (2016). Root-n consistent estimations of time dummies for the dynamic fixed effects logit models: Monte carlo illustrations. Technical report.

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.

Muris, C., Raposo, P., and Vandoros, S. (2020). A dynamic ordered logit model with fixed effects. *arXiv preprint arXiv:2008.05517*.

Narendranthan, W., Nickell, S., and Metcalf, D. (1985). An investigation into the incidence and dynamic structure of sickness and unemployment in britain, 1965–75. *Journal of the Royal Statistical Society: Series A (General)*, 148(3):254–267.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society*, pages 1–32.

Rasch, G. (1960). Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.

Appendix

A Partial Fraction Decomposition

Lemma 9. *For any reals* $u_1, u_2, \ldots, u_K$, $v_1, v_2, \ldots, v_K$ *and* $a_1, a_2, \ldots, a_K$, $K \geq 1$ *we have*

$$\frac{1}{1+\sum\limits_{k=1}^{K} e^{v_k+a_k}} + \sum_{k=1}^{K}(1-e^{u_k-v_k})\frac{e^{v_k+a_k}}{\left(1+\sum\limits_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} = \frac{1}{1+\sum\limits_{k=1}^{K} e^{u_k+a_k}}$$

and

$$\frac{e^{v_j+a_j}}{1+\sum\limits_{k=1}^{K} e^{v_k+a_k}} + (1-e^{-u_j+v_j})\frac{e^{u_j+a_j}}{\left(1+\sum\limits_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum\limits_{k=1}^{K} e^{u_k+a_k}\right)}+$$
$$\sum_{\substack{k=1\\k\neq j}}^{K}(1-e^{(u_k-u_j)-(v_k-v_j)})\frac{e^{v_k+a_k+u_j+a_j}}{\left(1+\sum\limits_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum\limits_{k=1}^{K} e^{u_k+a_k}\right)} = \frac{e^{u_j+a_j}}{1+\sum\limits_{k=1}^{K} e^{u_k+a_k}}$$

Proof.

$$\frac{1}{1+\sum_{k=1}^{K} e^{v_k+a_k}} + \sum_{k=1}^{K}(1-e^{u_k-v_k})\frac{e^{v_k+a_k}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} =$$
$$\frac{1+\sum_{k=1}^{K} e^{u_k+a_k}+\sum_{k=1}^{K} e^{v_k+a_k}-\sum_{k=1}^{K} e^{u_k+a_k}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)}$$
$$= \frac{1+\sum_{k=1}^{K} e^{v_k+a_k}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)}$$
$$= \frac{1}{1+\sum_{k=1}^{K} e^{u_k+a_k}}$$

and

$$
\begin{aligned}
&\frac{e^{v_j+a_j}}{1+\sum_{k=1}^{K} e^{v_k+a_k}} + (1-e^{-u_j+v_j})\frac{e^{u_j+a_j}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} + \\
&\sum_{\substack{k=1\\k\neq j}}^{K}(1-e^{(u_k-u_j)-(v_k-v_j)})\frac{e^{v_k+a_k+u_j+a_j}}{\left(1+\sum\limits_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} = \\
&\frac{e^{v_j+a_j}+\sum_{k=1}^{K} e^{v_j+a_j+u_k+a_k}+e^{u_j+a_j}-e^{v_j+a_j}+\sum\limits_{\substack{k=1\\k\neq j}}^{K} e^{v_k+a_k+u_j+a_j}-\sum_{\substack{k=1\\k\neq j}}^{K} e^{v_j+a_j+u_k+a_k}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} \\
&= \frac{e^{u_j+a_j}+e^{v_j+a_j+u_j+a_j}+\sum\limits_{\substack{k=1\\k\neq j}}^{K} e^{v_k+a_k+u_j+a_j}}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} \\
&= \frac{e^{u_j+a_j}\left(1+\sum\limits_{k=1}^{K} e^{v_k+a_k}\right)}{\left(1+\sum_{k=1}^{K} e^{v_k+a_k}\right)\left(1+\sum_{k=1}^{K} e^{u_k+a_k}\right)} \\
&= \frac{e^{u_j+a_j}}{1+\sum_{k=1}^{K} e^{u_k+a_k}}
\end{aligned}
$$

□

Lemma 10. *Fix $M \geq 2$, let $\mathcal{Y} = \{0,1\}^M$. Then, for any $k \in \mathcal{Y}$ and any reals $u_1, u_2, \ldots, u_M$, $v_1, v_2, \ldots, v_M$ and $a_1, a_2, \ldots, a_M$, we have*

$$
\prod_{m=1}^{M} \frac{e^{k_m(v_m+a_m)}}{1+e^{v_m+a_m}} + \sum_{l\in\mathcal{Y}\setminus\{k\}} \left[1-e^{\sum_{j=1}^{M}(l_j-k_j)(u_j-v_j)}\right] \prod_{m=1}^{M} \frac{e^{k_m(u_m+a_m)}}{1+e^{u_m+a_m}} \frac{e^{l_m(v_m+a_m)}}{1+e^{v_m+a_m}} = \prod_{m=1}^{M} \frac{e^{k_m(u_m+a_m)}}{1+e^{u_m+a_m}}
$$

Proof. Let

$$
LHS = \prod_{m=1}^{M} \frac{e^{k_m(v_m+a_m)}}{1+e^{v_m+a_m}} + \sum_{l\in\mathcal{Y}\setminus\{k\}} \left[1-e^{\sum_{j=1}^{M}(l_j-k_j)(u_j-v_j)}\right] \prod_{m=1}^{M} \frac{e^{k_m(u_m+a_m)}}{1+e^{u_m+a_m}} \frac{e^{l_m(v_m+a_m)}}{1+e^{v_m+a_m}}
$$

and let Num denote the numerator of LHS. We have:

$$
\begin{aligned}
Num &= Num_1 + Num_2 \\
Num_1 &= \prod_{m=1}^{M} e^{k_m(v_m+a_m)}(1+e^{u_m+a_m}) \\
Num_2 &= \sum_{l\in\mathcal{Y}\setminus\{k\}} \left[1 - e^{\sum_{j=1}^{M}(l_j-k_j)(u_j-v_j)}\right] \prod_{m=1}^{M} e^{k_m(u_m+a_m)+l_m(v_m+a_m)} \\
&= \prod_{m=1}^{M} e^{k_m(u_m+a_m)} \sum_{l\in\mathcal{Y}\setminus\{k\}} \prod_{m=1}^{M} e^{l_m(v_m+a_m)} - \sum_{l\in\mathcal{Y}\setminus\{k\}} e^{\sum_{j=1}^{M} l_j(u_j+a_j)+k_j(v_j+a_j)} \\
&= \prod_{m=1}^{M} e^{k_m(u_m+a_m)} \sum_{l\in\mathcal{Y}\setminus\{k\}} \prod_{m=1}^{M} e^{l_m(v_m+a_m)} - \prod_{m=1}^{M} e^{k_m(v_m+a_m)} \sum_{l\in\mathcal{Y}\setminus\{k\}} \prod_{m=1}^{M} e^{l_m(u_m+a_m)}
\end{aligned}
$$

Now, noting that

$$
\begin{aligned}
\sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} e^{l_m(v_m+a_m)} &= \prod_{m=1}^{M}(1+e^{v_m+a_m}) \\
\sum_{l\in\mathcal{Y}} \prod_{m=1}^{M} e^{l_m(u_m+a_m)} &= \prod_{m=1}^{M}(1+e^{u_m+a_m})
\end{aligned}
$$

we get

$$
\begin{aligned}
Num_2 &= \prod_{m=1}^{M} e^{k_m(u_m+a_m)} \sum_{l\in\mathcal{Y}\setminus\{k\}} \prod_{m=1}^{M} e^{l_m(v_m+a_m)} - \prod_{m=1}^{M} e^{k_m(v_m+a_m)} \sum_{l\in\mathcal{Y}\setminus\{k\}} \prod_{m=1}^{M} e^{l_m(u_m+a_m)} \\
&= \prod_{m=1}^{M} e^{k_m(u_m+a_m)} \left(\prod_{m=1}^{M}(1+e^{v_m+a_m}) - \prod_{m=1}^{M} e^{k_m(v_m+a_m)}\right) \\
&- \prod_{m=1}^{M} e^{k_m(v_m+a_m)} \left(\prod_{m=1}^{M}(1+e^{u_m+a_m}) - \prod_{m=1}^{M} e^{k_m(u_m+a_m)}\right) \\
&= \prod_{m=1}^{M} e^{k_m(u_m+a_m)}(1+e^{v_m+a_m}) - \prod_{m=1}^{M} e^{k_m(v_m+a_m)}(1+e^{u_m+a_m}) \\
&= \prod_{m=1}^{M} e^{k_m(u_m+a_m)}(1+e^{v_m+a_m}) - Num_1
\end{aligned}
$$

It follows that $Num = \prod_{m=1}^{M} e^{k_m(u_m+a_m)}(1+e^{v_m+a_m})$ and consequently

$$
LHS = \frac{\prod_{m=1}^{M} e^{k_m(u_m+a_m)}(1+e^{v_m+a_m})}{\prod_{m=1}^{M}(1+e^{u_m+a_m})(1+e^{v_m+a_m})} = \prod_{m=1}^{M} \frac{e^{k_m(u_m+a_m)}}{1+e^{u_m+a_m}}
$$

□

B Proofs of Propositions 1, 2, 3

Propositions 1, 2 and 3 all follow from the same strategy proof based on the the law of iterated expectations. We focus on Proposition 1 here and leave the other cases to the reader.

Take any t, s verifying $T-1 \geq t > s \geq 1$. For any $k \in \mathcal{Y}$, we have

$$\begin{aligned}
E\left[\psi_{\theta_0}^{k|k}(Y_{it-1}^{t+1}, Y_{is-1}^{s+1})|Y_{i0}, Y_{i1}^{s-1}, A_i\right] &= E\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}) - \phi_{\theta_0}^{k|k}(Y_{is+1}, Y_{is}, Y_{is-1})|Y_{i0}, Y_{i1}^{s-1}, A_i\right] \\
&= E\left[E\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1})|Y_{i0}, Y_{i1}^{t-1}, A_i\right]|Y_{i0}, Y_{i1}^{s-1}, A_i\right] - \pi^{k|k}(A_i) \\
&= E\left[\pi^{k|k}(A_i)|Y_{i0}, Y_{i1}^{s-1}, A_i\right] - \pi^{k|k}(A_i) \\
&= \pi^{k|k}(A_i) - \pi^{k|k}(A_i) \\
&= 0
\end{aligned}$$

The second and third equalities follow from the law of iterated expectation and Lemma 1.

C Proof of Lemma 2

The functional form proposed for the transition function $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq 0$. Hence

$$E\left[\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \frac{1}{1+e^{\gamma_0 Y_{it-1} + X_{it}'\beta_0 + A_i}} \times \left(\frac{e^{X_{it+1}'\beta_0 + A_i}}{1+e^{X_{it+1}'\beta_0 + A_i}}\phi_\theta^{0|0}(1, 0, Y_{it-1}, X_i) + \frac{1}{1+e^{X_{it+1}'\beta_0 + A_i}}\phi_\theta^{0|0}(0, 0, Y_{it-1}, X_i)\right)$$

Thus, to obtain the transition probability $\pi_t^{0|0}(A_i, X_i) = \frac{1}{1+e^{X_{it+1}'\beta_0 + A_i}}$ at $\theta = \theta_0$, we must set:

$$\begin{aligned}
\phi_\theta^{0|0}(1, 0, Y_{it-1}, X_i) &= e^{\gamma Y_{it-1} + (X_{it} - X_{it+1})'\beta} \\
\phi_\theta^{0|0}(0, 0, Y_{it-1}, X_i) &= 1 \\
\phi_\theta^{0|0}(k, 1, Y_{it-1}, X_i) &= 0, \quad \forall k \in \mathcal{Y}
\end{aligned}$$

This can be expressed compactly as: $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = (1 - Y_{it})e^{\gamma Y_{it+1}\left(Y_{it-1} - \Delta X_{it+1}'\beta\right)}$
Likewise, for $\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ we have:

$$E\left[\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \frac{e^{\gamma_0 Y_{it-1} + X_{it}'\beta_0 + A_i}}{1+e^{\gamma_0 Y_{it-1} + X_{it}'\beta_0 + A_i}} \times \left(\frac{e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}{1+e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}\phi_\theta^{1|1}(1, 1, Y_{it-1}, X_i) + \frac{1}{1+e^{\gamma_0 + X_{it+1}'\beta_0 + A_i}}\phi_\theta^{1|1}(0, 1, Y_{it-1}, X_i)\right)$$

Hence, to get $\pi_t^{1|1}(A_i, X_i) = \frac{e^{\gamma_0 + X'_{it+1}\beta_0 + A_i}}{1+e^{\gamma_0 + X'_{it+1}\beta_0 + A_i}}$ at $\theta = \theta_0$, we must set:

$$\phi_\theta^{1|1}(1, 1, Y_{it-1}, X_i) = 1$$
$$\phi_\theta^{1|1}(0, 1, Y_{it-1}, X_i) = e^{\gamma(1-Y_{it-1})+(X_{it+1}-X_{it})'\beta}$$
$$\phi_\theta^{1|1}(k, 0, Y_{it-1}, X_i) = 0, \quad \forall k \in \mathcal{Y}$$

This can be written succinctly as: $\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = Y_{it} e^{(1-Y_{it+1})\left(\gamma(1-Y_{it-1})+\beta \Delta X_{it+1}\right)}$

D Proof of Lemma 3

The result hinges on the simple identity provided in Lemma 9 that for any three reals v, u, a, we have:

$$\frac{1}{1+e^{v+a}} + (1-e^{u-v})\frac{e^{v+a}}{(1+e^{v+a})(1+e^{u+a})} = \frac{1}{(1+e^{u+a})}$$
$$\frac{e^{v+a}}{1+e^{v+a}} + (1-e^{-(u-v)})\frac{e^{u+a}}{(1+e^{v+a})(1+e^{u+a})} = \frac{e^{u+a}}{(1+e^{u+a})}$$

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

$$E\left[\zeta_{\theta_0}^{0|0}(Y_{it-1}^{t+1}, Y_{is-1}^{s}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] = E\left[(1-Y_{is}) + \omega_{t,s}^{0|0}(\theta_0) Y_{is} \phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right]$$
$$= \frac{1}{1+e^{\mu_s(\theta_0)+A_i}} + \omega_{t,s}^{0|0}(\theta_0) E\left[Y_{is} E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right]|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right]$$
$$= \frac{1}{1+e^{\mu_s(\theta_0)+A_i}} + \omega_{t,s}^{0|0}(\theta_0) E\left[Y_{is}|Y_{i0}, Y_{i1}^{s-1}, X_i, A_i\right] \frac{1}{1+e^{\kappa_t^{0|0}(\theta_0)+A_i}}$$
$$= \frac{1}{1+e^{\mu_s(\theta_0)+A_i}} + (1-e^{\kappa_t^{0|0}(\theta_0)-\mu_s(\theta_0)})\frac{e^{\mu_s(\theta_0)+A_i}}{(1+e^{\mu_s(\theta_0)+A_i})(1+e^{\kappa_t^{0|0}(\theta_0)+A_i})}$$
$$= \frac{1}{1+e^{\kappa_t^{0|0}(\theta_0)+A_i}}$$
$$= \pi_t^{0|0}(A_i, X_i)$$

The second equality follows from the measureability of the weight $\omega_{t,s}^{0|0}(\theta_0)$ with respect to the conditioning set. The third equality follows from the law of iterated expectations and Lemma 2. The penultimate equality uses the first mathematical identity presented above.

Similarly,

$$
\begin{aligned}
&E\left[\zeta^{1|1}_{\theta_0}(Y^{t+1}_{it-1},Y^{s}_{is-1},X_i)|Y_{i0},Y^{s-1}_{i1},X_i,A_i\right]=E\left[Y_{is}+\omega^{1|1}_{t,s}(\theta_0)(1-Y_{is})\phi^{1|1}_{\theta_0}(Y_{it+1},Y_{it},Y_{it-1},X_i)|Y_{i0},Y^{s-1}_{i1},X_i,A_i\right]\\
&=\frac{e^{\mu_s(\theta_0)+A_i}}{1+e^{\mu_s(\theta_0)+A_i}}+\omega^{1|1}_{t,s}(\theta_0)E\left[(1-Y_{is})E\left[\phi^{1|1}_{\theta_0}(Y_{it+1},Y_{it},Y_{it-1},X_i)|Y_{i0},Y^{t-1}_{i1},X_i,A_i\right]|Y_{i0},Y^{s-1}_{i1},X_i,A_i\right]\\
&=\frac{e^{\mu_s(\theta_0)+A_i}}{1+e^{\mu_s(\theta_0)+A_i}}+\omega^{1|1}_{t,s}(\theta_0)E\left[(1-Y_{is})|Y_{i0},Y^{s-1}_{i1},X_i,A_i\right]\frac{e^{\kappa^{1|1}_t(\theta_0)+A_i}}{1+e^{\kappa^{1|1}_t(\theta_0)+A_i}}\\
&=\frac{e^{\mu_s(\theta_0)+A_i}}{1+e^{\mu_s(\theta_0)+A_i}}+\left(1-e^{-(\kappa^{1|1}_t(\theta_0)-\mu_s(\theta_0))}\right)\frac{e^{\kappa^{1|1}_t(\theta_0)+A_i}}{(1+e^{\mu_s(\theta_0)+A_i})(1+e^{\kappa^{1|1}_t(\theta_0)+A_i})}\\
&=\frac{e^{\kappa^{1|1}_t(\theta_0)+A_i}}{1+e^{\kappa^{1|1}_t(\theta_0)+A_i}}\\
&=\pi^{1|1}_t(A_i,X_i)
\end{aligned}
$$

The second equality follows from the measureability of the weight $\omega^{0|0}_{t,s}(\theta_0)$ with respect to the conditioning set. The third equality follows from the law of iterated expectations and Lemma 2. The penultimate equality uses the second mathematical identity presented above.

E Connection to Kitazawa and Honoré-Weidner

Recall from Proposition 2 that when $T\geq 3$, our simplest moment conditions for t,s such that $T-1\geq t>s\geq 1$ write:

$$
\begin{aligned}
\psi^{0|0}_{\theta}(Y^{t+1}_{it-1},Y^{s}_{is-1},X_i)&=\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-\zeta^{0|0}_{\theta}(Y^{t+1}_{it-1},Y^{s}_{is-1},X_i)\\
&=\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-(1-Y_{is})-\omega^{0|0}_{t,s}(\theta)Y_{is}\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)\\
\psi^{1|1}_{\theta}(Y^{t+1}_{it-1},Y^{s}_{is-1},X_i)&=\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-\zeta^{1|1}_{\theta}(Y^{t+1}_{it-1},Y^{s}_{is-1},X_i)\\
&=\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-Y_{is}-\omega^{1|1}_{t,s}(\theta)(1-Y_{is})\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)
\end{aligned}
$$

where we know from Lemma 3 that

$$
\begin{aligned}
\omega^{0|0}_{t,s}(\theta)&=1-e^{(\kappa^{0|0}_t(\theta)-\mu_s(\theta))}\\
&=1-e^{(X_{it+1}-X_{is})'\beta-\gamma Y_{is-1}}\\
\omega^{1|1}_{t,s}(\theta)&=1-e^{-(\kappa^{1|1}_t(\theta)-\mu_s(\theta))}\\
&=1-e^{-\gamma(1-Y_{is-1})-(X_{it+1}-X_{is})'\beta)}
\end{aligned}
$$

Now, note that:

$$\tanh\left(\frac{\gamma(1-Y_{it-2})+(\Delta X_{it}+\Delta X_{it+1})'\beta}{2}\right)=\frac{1-e^{-\left(\gamma(1-Y_{it-2})+(\Delta X_{it}+\Delta X_{it+1})'\beta\right)}}{1+e^{-\left(\gamma(1-Y_{it-2})+(\Delta X_{it}+\Delta X_{it+1})'\beta\right)}}=\frac{\omega^{1|1}_{t,t-1}(\theta)}{2-\omega^{1|1}_{t,t-1}(\theta)}$$

$$\tanh\left(\frac{-\gamma Y_{it-2}+(\Delta X_{it}+\Delta X_{it+1})'\beta}{2}\right)=\frac{e^{-\gamma Y_{it-2}+(\Delta X_{it}+\Delta X_{it+1})'\beta}-1}{e^{-\gamma Y_{it-2}+(\Delta X_{it}+\Delta X_{it+1})'\beta}+1}=-\frac{\omega^{0|0}_{t,t-1}(\theta)}{2-\omega^{0|0}_{t,t-1}(\theta)}$$

and $\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)=\Upsilon_{it}$ and $1-\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)=U_{it}$. Thus, we have:

$$\begin{aligned}
(2-\omega^{1|1}_{t,t-1}(\theta))\hbar U_{it}&=(2-\omega^{1|1}_{t,t-1}(\theta))(U_{it}-Y_{it-1})+\omega^{0|0}_{t,t-1}(\theta)\left(U_{it}+Y_{it-1}-2U_{it}Y_{it-1}\right)\\
&=2\left[U_{it}-Y_{it-1}+\omega^{0|0}_{t,t-1}(\theta)Y_{it-1}(1-U_{it})\right]\\
&=2\left[1-\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-Y_{it-1}+\omega^{0|0}_{t,t-1}(\theta)Y_{it-1}\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)\right]\\
&=-2\left[\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-(1-Y_{it-1})-\omega^{0|0}_{t,t-1}(\theta)Y_{it-1}\phi^{0|0}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)\right]\\
&=-2\psi^{0|0}_{\theta}(Y^{t+1}_{it-1},Y^{t-1}_{it-2},X_i)\\
(2-\omega^{1|1}_{t,t-1}(\theta))\hbar\Upsilon_{it}&=(2-\omega^{1|1}_{t,t-1}(\theta))(\Upsilon_{it}-Y_{it-1})-\omega^{1|1}_{t,t-1}(\theta)\left(\Upsilon_{it}+Y_{it-1}-2\Upsilon_{it}Y_{it-1}\right)\\
&=2\left[\Upsilon_{it}-Y_{it-1}-\omega^{1|1}_{t,t-1}(\theta)\Upsilon_{it}\left(1-Y_{it-1}\right)\right]\\
&=2\left[\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)-Y_{it-1}-\omega^{1|1}_{t,t-1}(\theta)\phi^{1|1}_{\theta}(Y_{it+1},Y_{it},Y_{it-1},X_i)\left(1-Y_{it-1}\right)\right]\\
&=2\psi^{1|1}_{\theta}(Y^{t+1}_{it-1},Y^{t-1}_{it-2},X_i)
\end{aligned}$$

To establish the connection to the work of Honoré and Kyriazidou (2000), it is useful to re-write the moment functions slightly differently. By re-arranging terms, one obtains the following for $T=3$

$$\begin{aligned}
\psi^{0|0}_{\theta}(Y^3_1,Y^1_{i0},X_i)&=(1-Y_{i1})\phi^{0|0}_{\theta}(Y^3_{i1},X_i)+e^{(X_{i3}-X_{i1})'\beta-\gamma Y_{i0}}Y_{i1}\phi^{0|0}_{\theta}(Y^3_{i1},X_i)-(1-Y_{i1})\\
&=e^{(X_{i2}-X_{i3})'\beta}(1-Y_{i1})(1-Y_{i2})Y_{i3}+(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})\\
&+e^{(X_{i2}-X_{i1})'\beta+\gamma(1-Y_{i0})}Y_{i1}(1-Y_{i2})Y_{i3}\\
&+e^{(X_{i3}-X_{i1})'\beta-\gamma Y_{i0}}Y_{i1}(1-Y_{i2})(1-Y_{i3})\\
&-(1-Y_{i1})\\
&=(e^{(X_{i2}-X_{i3})'\beta}-1)(1-Y_{i1})(1-Y_{i2})Y_{i3}\\
&+e^{(X_{i2}-X_{i1})'\beta+\gamma(1-Y_{i0})}Y_{i1}(1-Y_{i2})Y_{i3}\\
&+e^{(X_{i3}-X_{i1})'\beta-\gamma Y_{i0}}Y_{i1}(1-Y_{i2})(1-Y_{i3})\\
&-(1-Y_{i1})Y_{i2}
\end{aligned}\tag{13}$$

where the last line uses the fact that: $(1-Y_{i1})=(1-Y_{i1})Y_{i2}+(1-Y_{i1})(1-Y_{i2})Y_{i3}+(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})$ to make some cancellations. For the initial condition, $Y_{i0}=0$, equation (13) corresponds to their moment function m_0^b which they express in an extensive form. For $Y_{i0} = 1$, we get instead m_1^b. Similarly,

$$\begin{aligned}
\psi_\theta^{1|1}(Y_{i1}^3, Y_{i0}^1, X_i) &= Y_{i1}\phi_\theta^{1|1}(Y_{i1}^3, X_i) + e^{-\gamma(1-Y_{i0})-(X_{i3}-X_{i1})'\beta)}(1-Y_{i1})\phi_\theta^{1|1}(Y_{i1}^3, X_i) - Y_{i1} \\
&= e^{(X_{i3}-X_{i2})'\beta}Y_{i1}Y_{i2}(1-Y_{i3}) + Y_{i1}Y_{i2}Y_{i3} \\
&+ e^{(X_{i1}-X_{i2})'\beta+\gamma Y_{i0}}(1-Y_{i1})Y_{i2}(1-Y_{i3}) \\
&+ e^{(X_{i1}-X_{i3})'\beta-\gamma(1-Y_{i0})}(1-Y_{i1})Y_{i2}Y_{i3} \\
&- Y_{i1} \\
&= (e^{(X_{i3}-X_{i2})'\beta} - 1)Y_{i1}Y_{i2}(1-Y_{i3}) \\
&+ e^{(X_{i1}-X_{i2})'\beta+\gamma Y_{i0}}(1-Y_{i1})Y_{i2}(1-Y_{i3}) \\
&+ e^{(X_{i1}-X_{i3})'\beta-\gamma(1-Y_{i0})}(1-Y_{i1})Y_{i2}Y_{i3} \\
&- Y_{i1}(1-Y_{i2})
\end{aligned} \tag{14}$$

where the last line uses the fact that: $Y_{i1} = Y_{i1}(1-Y_{i2}) + Y_{i1}Y_{i2}Y_{i3} + Y_{i1}Y_{i2}(1-Y_{i3})$. For the initial condition $Y_{i0} = 0$, equation (14) gives their moment function m_0^a and for $Y_{i0} = 1$, we get m_1^a. Our moments are thus identical, at least for the case $T = 3$.

F Proof of Theorem 1

We start by proving the following Lemma

Lemma 11. *In model (9), with $T \geq 2$ and $t \in \{1, \ldots, T-1\}$, let*

$$\begin{aligned}
\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i) &= (1-Y_{it})e^{Y_{it+1}(\gamma_1 Y_{it-1} - \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} - \Delta X_{it+1}'\beta)} \\
\phi_\theta^{1|1}(Y_{it+1}, Y_t, Y_{it-p}^{t-1}, X_i) &= Y_{it}e^{(1-Y_{it+1})\left(\gamma_1(1-Y_{it-1}) + \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} + \Delta X_{it+1}'\beta\right)}
\end{aligned}$$

Then,

$$\begin{aligned}
E\left[\phi_{\theta_0}^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-1}, X_i, A_i\right] &= \pi_t^{0|0,Y_{it-(p-1)}^{t-1}}(A_i, X_i) = \frac{1}{1 + e^{\sum_{l=2}^p \gamma_{0l} Y_{it+1-l} + X_{it+1}'\beta_0 + A_i}} \\
E\left[\phi_{\theta_0}^{1|1}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-1}, X_i, A_i\right] &= \pi_t^{1|1,Y_{it-(p-1)}^{t-1}}(A_i, X_i) = \frac{e^{\gamma_{01} + \sum_{l=2}^p \gamma_{0l} Y_{it+1-l} + X_{it+1}'\beta_0 + A_i}}{1 + e^{\gamma_{01} + \sum_{l=2}^p \gamma_{0l} Y_{it+1-l} + X_{it+1}'\beta_0 + A_i}}
\end{aligned}$$

Instead of verifying the result directly from the expression given in the Lemma, it is easier to start from the heuristic idea, emphasized throughout the text, that we look for two functions such that:

$$\begin{aligned}
&\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i) = (1-Y_{it})\phi_\theta^{0|0}(Y_{it+1}, 0, Y_{it-p}^{t-1}, X_i) \\
&\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = Y_{it}\phi_\theta^{1|1}(Y_{it+1}, 1, Y_{it-p}^{t-1}, X_i) \\
&E\left[\phi_{\theta_0}^{k|k}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-1}, X_i, A_i\right] = \pi_t^{k|k,Y_{it-(p-1)}^{t-1}}(A_i, X_i), \quad \forall k \in \mathcal{Y}
\end{aligned}$$

By definition, $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)$ is null when $Y_{it} \neq 0$. Hence

$$E\left[\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-1}, X, A\right] = \frac{1}{1+e^{\sum_{l=1}^p \gamma_{0l} Y_{it-l}+X_{it}'\beta_0+A_i}} \times ($$
$$\frac{e^{\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}{1+e^{\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}\phi_\theta^{0|0}(1, 0, Y_{it-p}^{t-1}, X_i) + \frac{1}{1+e^{\gamma_{02} Y_{it-1}+X_{it+1}'\beta_0+A_i}}\phi_\theta^{0|0}(0, 0, Y_{it-p}^{t-1}, X_i))$$

Thus, to obtain the transition probability $\pi_t^{0|0, Y_{it-(p-1)}^{t-1}}(A_i, X_i) = \frac{1}{1+e^{\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}$ at $\theta = \theta_0$, we must set:

$$\phi_\theta^{0|0}(1, 0, Y_{it-p}^{t-1}, X_i) = e^{\gamma_1 Y_{it-1} - \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} - \Delta X_{it+1}'\beta}$$
$$\phi_\theta^{0|0}(0, 0, Y_{it-p}^{t-1}, X_i) = 1$$
$$\phi_\theta^{0|0}(k, 1, Y_{it-p}^{t-1}, X_i) = 0, \forall k \in \mathcal{Y}$$

more compactly this writes, $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i) = (1-Y_{it})e^{Y_{it+1}(\gamma_1 Y_{it-1} - \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} - \Delta X_{it+1}'\beta)}$. Analogously, $\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i)$ is null when $Y_{it} \neq 1$. Hence

$$E\left[\phi_\theta^{1|1}(Y_{it+1}, Y_{it}, Y_{it-p}^{t-1}, X_i))|Y_i^0, Y_1^{t-1}, X, A\right] = \frac{e^{\sum_{l=1}^p \gamma_{0l} Y_{it-l}+X_{it}'\beta_0+A_i}}{1+e^{\sum_{l=1}^p \gamma_{0l} Y_{it-l}+X_{it}'\beta_0+A_i}} \times ($$
$$\frac{e^{\gamma_{01}+\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}{1+e^{\gamma_{01}+\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}\phi_\theta^{1|1}(1, 1, Y_{it-p}^{t-1}, X_i) + \frac{1}{1+e^{\gamma_{01}+\gamma_{02} Y_{it-1}+X_{it+1}'\beta_0+A_i}}\phi_\theta^{1|1}(0, 1, Y_{it-p}^{t-1}, X_i))$$

Consequently, to get $\pi_t^{1|1, Y_{it-(p-1)}^{t-1}}(A_i, X_i) = \frac{e^{\gamma_{01}+\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}{1+e^{\gamma_{01}+\sum_{l=2}^p \gamma_{0l} Y_{it+1-l}+X_{it+1}'\beta_0+A_i}}$ at $\theta = \theta_0$, we must set:

$$\phi_\theta^{1|1}(1, 1, Y_{it-p}^{t-1}, X_i) = 1$$
$$\phi_\theta^{1|1}(0, 1, Y_{it-p}^{t-1}, X_i) = e^{\gamma_1(1-Y_{it-1}) + \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} + \Delta X_{it+1}'\beta}$$
$$\phi_\theta^{1|1}(k, 0, Y_{it-p}^{t-1}, X_i) = 0, \forall k \in \mathcal{Y}$$

This can be written succinctly as: $\phi_\theta^{1|1}(Y_{it+1}, Y_t, Y_{it-p}^{t-1}, X_i) = Y_{it}e^{(1-Y_{it+1})\left(\gamma_1(1-Y_{it-1}) + \sum_{l=2}^p \gamma_l \Delta Y_{it+1-l} + \Delta X_{it+1}'\beta\right)}$. This completes the proof of the Lemma.

Now, for $T \geq p+1$ fix $t \in \{p, \ldots, T-1\}$ and $y = (y_1, \ldots, y_p) = y_1^p \in \{0,1\}^p$. We will prove by finite induction the statement $\mathcal{P}(k)$:

$$E\left[\phi_{\theta_0}^{y_1|y_1^{k+1}}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right] = \pi_t^{y_1|y_1^{k+1}, Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i)$$

for $k = 0, \ldots, p-2$ for $p \geq 2$.

Base step:

$\mathcal{P}(0)$ is true by Lemma 11 which also deals with the edge case $p = 2$. Thus, let us assume $p \geq 3$ in the remainder of the induction argument.

Induction Step:

Suppose $\mathcal{P}(k)$ is true for some $k \in \{0, \ldots, p-3\}$, we show that $\mathcal{P}(k+1)$ is true. Using the law of iterated expectations, the induction hypothesis $\mathcal{P}(k)$ and the identities of Lemma 9, we have:

If $y_1 = 0, y_{k+1} = 1$

$$E\left[\phi_{\theta_0}^{0|0,y_2^k,1}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= E\left[(1 - Y_{it-k}) + w_t^{0|0,y_2^k,1}(\theta_0)\phi_{\theta_0}^{0|0,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{1}{1 + e^{u_{t-k}(\theta_0)+A_i}} + w_t^{0|0,y_2^k,1}(\theta_0)E\left[E\left[\phi_{\theta_0}^{0|0,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-k}, X_i, A_i\right] Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{1}{1 + e^{u_{t-k}(\theta_0)+A_i}} w_t^{0|0,y_2^k,1}(\theta_0)E\left[\pi_t^{0|0,y_2^k,Y_{it-(p-1)}^{t-k}}(A_i, X_i)Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{1}{1 + e^{u_{t-k}(\theta_0)+A_i}} + w_t^{0|0,y_2^k,1}(\theta_0)E\left[\frac{1}{1 + e^{\sum_{r=2}^k \gamma_{0r}y_r + \sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)} + X_{it+1}'\beta_0 + A_i}} Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{1}{1 + e^{u_{t-k}(\theta_0)+A_i}} + (1 - e^{(k_t^{0|0,y_2^k,1}(\theta_0) - u_{t-k}(\theta_0))})\frac{1}{1 + e^{k_t^{0|0,y_2^k,1}(\theta_0)+A_i}}\frac{e^{u_{t-k}(\theta_0)+A_i}}{1 + e^{u_{t-k}(\theta_0)+A_i}}$$
$$= \frac{1}{1 + e^{k_t^{0|0,y_2^k,1}(\theta_0)+A_i}}$$
$$= \pi_t^{0|0,y_2^k,1,Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i)$$

If $y_1 = 0, y_{k+1} = 0$

$$E\left[\phi_{\theta_0}^{0|0,y_2^k,0}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= E\left[1 - Y_{it-k} - w_t^{0|0,y_2^k,0}(\theta_0)\left(1 - \phi_{\theta_0}^{0|0,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)\right)(1 - Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \frac{e^{u_{t-k}(\theta_0)+A_i}}{1 + e^{u_{t-k}(\theta_0)+A_i}}$$
$$- w_t^{0|0,y_2^k,0}(\theta_0)E\left[E\left[\left(1 - \phi_{\theta_0}^{0|0,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)\right)|Y_i^0, Y_{i1}^{t-k}, X_i, A_i\right](1 - Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \frac{e^{u_{t-k}(\theta_0)+A_i}}{1 + e^{u_{t-k}(\theta_0)+A_i}} - w_t^{0|0,y_2^k,0}(\theta_0)E\left[(1 - \pi_t^{0|0,y_2^k,Y_{it-(p-1)}^{t-k}}(A_i, X_i))(1 - Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \frac{e^{u_{t-k}(\theta_0)+A_i}}{1 + e^{u_{t-k}(\theta_0)+A_i}} - w_t^{0|0,y_2^k,0}(\theta_0)E\left[\frac{e^{\sum_{r=2}^k \gamma_{0r}y_r + \sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)} + X_{it+1}'\beta_0 + A_i}}{1 + e^{\sum_{r=2}^k \gamma_{0r}y_r + \sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)} + X_{it+1}'\beta_0 + A_i}}(1 - Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \left(\frac{e^{u_{t-k}(\theta_0)+A_i}}{1 + e^{u_{t-k}(\theta_0)+A_i}} + (1 - e^{-(k_t^{0|0,y_2^k,0}(\theta_0) - u_{t-k}(\theta_0))})\frac{e^{k_t^{0|0,y_2^k,0}(\theta_0)+A_i}}{1 + e^{k_t^{0|0,y_2^k,0}(\theta_0)+A_i}}\frac{1}{1 + e^{u_{t-k}(\theta_0)+A_i}}\right)$$
$$= 1 - \frac{e^{k_t^{0|0,y_2^k,0}(\theta_0)+A_i}}{1 + e^{k_t^{0|0,y_2^k,0}(\theta_0)+A_i}}$$
$$= \frac{1}{1 + e^{k_t^{0|0,y_2^k,0}(\theta_0)+A_i}}$$
$$= \pi_t^{0|0,y_2^k,0,Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i)$$

If $y_1 = 1, y_{k+1} = 0$

$$E\left[\phi_{\theta_0}^{1|1,y_2^k,0}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= E\left[Y_{it-k} + w_t^{1|1,y_2^k,0}(\theta_0)\phi_{\theta_0}^{1|1,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)(1-Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{e^{u_{t-k}(\theta_0)+A_i}}{1+e^{u_{t-k}(\theta_0)+A_i}} + w_t^{1|1,y_2^k,0}(\theta_0)\times$$
$$E\left[E\left[\phi_{\theta_0}^{1|1,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-k}, X_i, A_i\right](1-Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{e^{u_{t-k}(\theta_0)+A_i}}{1+e^{u_{t-k}(\theta_0)+A_i}} + w_t^{1|1,y_2^k,0}(\theta_0)E\left[\pi_t^{1|1,y_2^k,Y_{it-(p-1)}^{t-k}}(A_i, X_i)(1-Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{e^{u_{t-k}(\theta_0)+A_i}}{1+e^{u_{t-k}(\theta_0)+A_i}} + w_t^{1|1,y_2^k,0}(\theta_0)E\left[\frac{e^{\gamma_{01}+\sum_{r=2}^k \gamma_{0r}y_r+\sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)}+X_{it+1}'\beta_0+A_i}}{1+e^{\gamma_{01}+\sum_{r=2}^k \gamma_{0r}y_r+\sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)}+X_{it+1}'\beta_0+A_i}}(1-Y_{it-k})|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= \frac{e^{u_{t-k}(\theta_0)+A_i}}{1+e^{u_{t-k}(\theta_0)+A_i}} + (1-e^{-(k_t^{1|1,y_2^k,0}(\theta_0)-u_{t-k}(\theta_0))})\frac{e^{k_t^{1|1,y_2^k,0}(\theta_0)+A_i}}{1+e^{k_t^{1|1,y_2^k,0}(\theta_0)+A_i}}\frac{1}{1+e^{u_{t-k}(\theta_0)+A_i}}$$
$$= \frac{e^{k_t^{1|1,y_2^k,0}(\theta_0)+A_i}}{1+e^{k_t^{1|1,y_2^k,0}(\theta_0)+A_i}}$$
$$= \pi_t^{1|1,y_2^k,0,Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i)$$

If $y_1 = 1, y_{k+1} = 1$

$$E\left[\phi_{\theta_0}^{1|1,y_2^k,1}(Y_{it+1}, Y_{it}, Y_{it-(p+k)}^{t-1}, X_i)|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= E\left[1-(1-Y_{it-k}) - w_t^{1|1,y_2^k,1}(\theta_0)\left(1-\phi_{\theta_0}^{1|1,y_2^k}(Y_{it+1}, Y_{it}, Y_{it-(p+k-1)}^{t-1}, X_i)\right)Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \frac{1}{1+e^{u_{t-k}(\theta_0)+A_i}}$$
$$- w_t^{1|1,y_2^k,1}(\theta_0)E\left[E\left[\left(1-\pi_t^{1|1,y_2^k,Y_{it-(p-1)}^{t-k}}(A_i, X_i)\right)|Y_i^0, Y_{i1}^{t-k}, X_i, A_i\right]Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \frac{1}{1+e^{u_{t-k}(\theta_0)+A_i}} - w_t^{1|1,y_2^k,1}(\theta_0)E\left[\frac{1}{1+e^{\gamma_{01}+\sum_{r=2}^k \gamma_{0r}y_r+\sum_{r=k+1}^p \gamma_{0r}Y_{it-(r-1)}+X_{it+1}'\beta_0+A_i}}Y_{it-k}|Y_i^0, Y_{i1}^{t-(k+1)}, X_i, A_i\right]$$
$$= 1 - \left(\frac{1}{1+e^{u_{t-k}(\theta_0)+A_i}} + (1-e^{(k_t^{1|1,y_2^k,1}(\theta_0)-u_{t-k}(\theta_0))})\frac{1}{1+e^{k_t^{1|1,y_2^k,1}(\theta_0)+A_i}}\frac{e^{u_{t-k}(\theta_0)+A_i}}{1+e^{u_{t-k}(\theta_0)+A_i}}\right)$$
$$= 1 - \frac{1}{1+e^{k_t^{1|1,y_2^k,1}(\theta_0)+A_i}}$$
$$= \frac{e^{k_t^{1|1,y_2^k,1}(\theta_0)+A_i}}{1+e^{k_t^{1|1,y_2^k,1}(\theta_0)+A_i}}$$
$$= \pi_t^{1|1,y_2^k,1,Y_{it-(p-1)}^{t-(k+1)}}(A_i, X_i)$$

Putting these intermediate results together, we have effectively proved that

$$E\left[\phi_{\theta_0}^{y_1|y_1^{k+1}}(Y_{it+1},Y_{it},Y_{it-(p+k)}^{t-1},X_i)|Y_i^0,Y_{i1}^{t-(k+1)},X_i,A_i\right]=\pi_t^{y_1|y_1^{k+1},Y_{it-(p-1)}^{t-(k+1)}}(A_i,X_i)$$

which completes the induction argument.

Now, it only remains to show that

$$E\left[\phi_{\theta_0}^{y_1|y_1^{p}}(Y_{it+1},Y_{it},Y_{it-(2p-1)}^{t-1},X_i)|Y_i^0,Y_{i1}^{t-p},X_i,A_i\right]=\pi_t^{y_1|y_1^{p}}(A_i,X_i)$$

To this end, it suffices to perform calculations identical to those used in the induction argument but using this time

$$\begin{aligned}
&E\left[\phi_{\theta_0}^{y_1|y_1^{p-1}}(Y_{it+1},Y_{it},Y_{it-(2p-2)}^{t-1},X_i)|Y_i^0,Y_{i1}^{t-(p-1)},X_i,A_i\right]=\pi_t^{y_1|y_1^{p-1},Y_{it-(p-1)}}(A_i,X_i)\\
&k_t^{y_1|y_1^p}(\theta)=\sum_{r=1}^{p}\gamma_r y_r+X_{it+1}'\beta\\
&u_{t-(p-1)}(\theta)=\sum_{r=1}^{p}\gamma_r Y_{it-(r+p-1)}+X_{it-(p-1)}'\beta\\
&w_t^{y_1|y_1^p}(\theta)=\left[1-e^{(k_t^{y_1|y_1^p}(\theta)-u_{t-(p-1)}(\theta))}\right]^{y_p}\left[1-e^{-(k_t^{y_1|y_1^p}(\theta)-u_{t-(p-1)}(\theta))}\right]^{1-y_p}
\end{aligned}$$

This concludes the proof of the theorem.

G Proof of Theorem 2

Without loss of generality and for notational simplicity, we will establish the result for $Y_i^0=0_p$. It will be clear from the argument below that nothing hinges on that particular case.

Our starting point is that equation (1) for a given T, initial condition y^0 and regressors x can be written equivalently as:

$$\sum_{(y_1,\ldots,y_T)\in\mathcal{Y}^T}\psi_{\theta_0}(y_1^T,y^0,x)P(Y_{i1}=y_1,\ldots,Y_{iT}=y_T|Y_i^0=y^0,X_i=x,A_i=a)=0,\quad\forall a\in\mathbb{R}$$

This formulation is interesting as it really clarifies that the existence of a valid moment function for a given T is equivalent to the existence of a linear relationship between the conditional probabilities of all choice histories of length T, $y_1^T=(y_1,\ldots,y_T)\in\mathcal{Y}^T$, given (y^0,x,a) when viewed as functions of a. We will now prove that such a linear dependence does not exist under our working assumptions for $T\leq p+1$. To this end, we will show via finite induction that the statement $\mathcal{P}(k)$:

"the conditional probabilities of all choice histories of length k, $y_1^k\in\mathcal{Y}^k$, given $Y_i^0=0_p,X_i=x,A_i$ are linearly independent as functions of A_i"

holds for all $k = 1, \ldots, p+1$.

Base step:

For $k = 1$, the two possible choice histories are $y_1 = 0$ and $y_1 = 1$ with conditional probabilities given by:

$$P(Y_{i1} = 1 | Y_i^0 = 0_p, X_i = x, A_i = a) = \frac{1}{1 + e^{x_1' \beta_0 + a}}$$
$$P(Y_{i1} = 0 | Y_i^0 = 0_p, X_i = x, A_i = a) = \frac{e^{x_1' \beta_0 + a}}{1 + e^{x_t' \beta_0 + a}}$$

Only one of the two probabilities has e^a in the numerator so the two are clearly linearly independent. [9]

Induction Step:

Suppose $\mathcal{P}(k)$ is true for some $k \in \{1, \ldots, p\}$, we show that $\mathcal{P}(k+1)$ is true.

Towards a contradiction, suppose that the 2^{k+1} possible choice histories have conditional probabilities that are linearly dependent. Then, this means that there exists a collection of scalars $(\lambda_{y_1, \ldots, y_{k+1}})_{y_1^{k+1} \in \{0,1\}^{k+1}}$ not all zeros such that:

$$\sum_{y_1^{k+1} \in \{0,1\}^{k+1}} \lambda_{y_1, \ldots, y_{k+1}} P(Y_{i1} = y_1, \ldots, Y_{ik+1} = y_{k+1} | Y_i^0 = 0_p, X_i = x, A_i = a) = 0, \quad \forall a \in \mathbb{R} \tag{15}$$

Let us use the following reparameterization $u = e^a$, $B_t = e^{x_t' \beta}$ for all $t \in \{1, \ldots, k+1\}$ and $C_t = e^{\gamma_t}$ for all $t \in \{1, \ldots, k\}$. Then (15), can be reformulated as

$$\sum_{y_1^k \in \{0,1\}^k} Q_{y_1, \ldots, y^k}(u) = 0, \quad \forall u \in \mathbb{R}_+ \tag{16}$$

where

$$\begin{aligned} Q_{y_1, \ldots, y^k}(u) = \lambda_{y_1, \ldots, y_k, 0} P(Y_{i1} = y_1, \ldots, Y_{ik} = y_k, Y_{ik+1} = 0 | Y_i^0 = 0_p, X_i = x, A_i = \ln u) \\ + \lambda_{y_1, \ldots, y_k, 1} P(Y_{i1} = y_1, \ldots, Y_{ik} = y_k, Y_{ik+1} = 1 | Y_i^0 = 0_p, X_i = x, A_i = \ln u) \end{aligned}$$

[9] $\{1, P(Y_1 = 1 | Y^0 = 0_p, X = x, A = a), P(Y_1 = 0 | Y^0 = 0_p, X = x, A = a)\}$ are not linearly independent however since the two probabilities sum to 1)

that is

$$
\begin{aligned}
Q_{y_1,\ldots,y^k}(u) &= \lambda_{y_1,\ldots,y_k,0}\frac{(B_1u)^{y_1}\prod_{t=2}^{k}\left(B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)^{y_t}}{(1+B_1u)\prod_{t=2}^{k+1}\left(1+B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)} + \lambda_{y_1,\ldots,y_k,1}\frac{(B_1u)^{y_1}\prod_{t=2}^{k}\left(B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)^{y_t}B_{k+1}\prod_{s=1}^{k}C_s^{y_{t-s}}u}{(1+B_1u)\prod_{t=2}^{k+1}\left(1+B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)} \\
&= \frac{(B_1u)^{y_1}\prod_{t=2}^{k}\left(B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)^{y_t}}{(1+B_1u)\prod_{t=2}^{k+1}\left(1+B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)}\left(\lambda_{y_1,\ldots,y_k,0}+\lambda_{y_1,\ldots,y_k,1}B_{k+1}\prod_{s=1}^{k}C_s^{y_{t-s}}u\right)
\end{aligned}
$$

Define $Q:\mathbb{R}\mapsto\mathbb{R}$ such that, $\forall u\in\mathbb{R},\quad Q(u)=\sum_{y_1^k\in\{0,1\}^k}Q_{y_1,\ldots,y^k}(u)$ where the definition of $Q_{y_1,\ldots,y^k}(u)$ shown in the last display is extended to the real line. We know that $Q(.)$ is null on $\mathbb{R}_+$ from (16). We argue that $Q(.)$ must be the null function on the entire real line. Indeed, if we define

$$
R(u)=(1+B_1u)\prod_{t=1}^{k}\prod_{y_1^t\in\mathcal{Y}^t}\left(1+B_{t+1}\prod_{s=1}^{t}C_s^{y_s}u\right)
$$

then, by standard properties of polynomials, $S(u)=Q(u)R(u)$ (which is nothing but the numerator of $Q(u)$ after putting everything to a common denominator) must be the null function on the real line given that it has an infinite amount of roots on $\mathbb{R}_+$. From inspection, it is clear that $R(u)$ does not vanish on $\mathbb{R}_+$ since its roots are negative given the definition of B_t and C_t. Therefore, it cannot be the null polynomial and hence it must be the case that $Q(u)=0$, for all $u\in\mathbb{R}$. In other words, (16) holds on the entire real line.

Now, a critical observation is that due to our assumptions on the common parameters and x, the term $Q_{y_1,\ldots,y^k}(u)$ is the only of the 2^k summand in $Q(u)$ featuring the term $T_{y_1,\ldots,y^k}(u)=\left(1+B_{k+1}\prod_{s=1}^{k}C_s^{y_{k+1-s}}u\right)$ in its denominator. The root of $T_{y_1,\ldots,y^k}(u)$ is $-\left(B_{k+1}\prod_{s=1}^{k}C_s^{y_{k+1-s}}\right)^{-1}$. It follows that if we consider the function $Q(u)T_{y_1,\ldots,y^k}(u)$ and evaluate it at $u=-\left(B_{k+1}\prod_{s=1}^{k}C_s^{y_{k+1-s}}\right)^{-1}$, one gets

$$
\lambda_{y_1,\ldots,y_k,0}-\lambda_{y_1,\ldots,y_k,1}B_{k+1}\prod_{s=1}^{k}C_s^{y_{k+1-s}}\left(B_{k+1}\prod_{s=1}^{k}C_s^{y_{k+1-s}}\right)^{-1}=\lambda_{y_1,\ldots,y_k,0}-\lambda_{y_1,\ldots,y_k,1}=0
$$

That is, $\lambda_{y_1,\ldots,y_k,0}=\lambda_{y_1,\ldots,y_k,1}=\lambda_{y_1,\ldots,y_k}$. An immediate consequence, is

$$
Q_{y_1,\ldots,y^k}(u)=\lambda_{y_1,\ldots,y_k}\frac{(B_1u)^{y_1}\prod_{t=2}^{k}\left(B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)^{y_t}}{(1+B_1u)\prod_{t=2}^{k}\left(1+B_t\prod_{s=1}^{t-1}C_s^{y_{t-s}}u\right)}
$$

which implies $Q_{y_1,\ldots,y^k}(e^a)=\lambda_{y_1,\ldots,y_k}P(Y_{i1}=y_1,\ldots,Y_{ik}=y_k|Y_i^0=0_p,X_i=x,A_i=a)$. Since this is true for all of the 2^k summands $Q_{y_1,\ldots,y^k}(e^a)$ of $Q(e^a)$, $Q(e^a)=0$ implies that that the conditional probabilities of all choice histories of length k, $y_1^k \in \mathcal{Y}^k$, given $Y_i^0 = y^0, X_i = x, A_i$ are linearly dependent as functions of A_i. This contradicts the induction hypothesis $\mathcal{P}(k)$. Therefore, $\mathcal{P}(k+1)$ is true which concludes the proof.

H Identification of the pure AR(2) logit model with $T = 3$

We have $T = p+1$ here since $p = 2$, so by Theorem 1, we know that we can compute the transition functions associated to: $\pi^{0|0,0}(A_i), \pi^{0|0,1}(A_i), \pi^{1|1,0}(A_i), \pi^{1|1,1}(A_i)$. They are given by:

$$\phi_\theta^{0|0}(Y_{i3}, Y_{i2}, Y_{i0}^1) = (1 - Y_{i2})e^{Y_{i3}(\gamma_{01}Y_{i1} - \gamma_{02}(Y_{i1} - Y_{i0}))}$$
$$\phi_\theta^{1|1}(Y_{i3}, Y_2, Y_{i0}^1) = Y_{i2}e^{(1-Y_{i3})\left(\gamma_{01}(1-Y_{i1}) + \gamma_{02}(Y_{i1} - Y_{i0})\right)}$$
$$\phi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} - (1 - e^{\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}})\left(1 - \phi_\theta^{0|0}(Y_{i3}, Y_{i2}, Y_{i0}^1, X_i)\right)(1 - Y_{i1})$$
$$\phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} + (1 - e^{\gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})\phi_\theta^{0|0}(Y_{i3}, Y_{i2}, Y_{i0}^1)Y_{i1}$$
$$\phi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = Y_{i1} + (1 - e^{-(\gamma_1 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}))})\phi_\theta^{1|1}(Y_{i3}, Y_{i2}, Y_{i0}^1, X_i)(1 - Y_{i1})$$
$$\phi_\theta^{1|1,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - (1 - Y_{i1}) - (1 - e^{\gamma_1 + \gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})\left(1 - \phi_\theta^{1|1}(Y_{i3}, Y_{i2}, Y_{i0}^1, X_i)\right)Y_{i1}$$

After simplifications:

$$\phi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} - (1 - e^{\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}})\left(1 - (1 - Y_{i2})e^{\gamma_{02} Y_{i3} Y_{i0}}\right)(1 - Y_{i1})$$
$$\phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} + (1 - e^{\gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})(1 - Y_{i2})e^{Y_{i3}(\gamma_{01} - \gamma_{02}(1 - Y_{i0}))}Y_{i1}$$
$$\phi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = Y_{i1} + (1 - e^{-(\gamma_1 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1}))})Y_{i2}e^{(1-Y_{i3})(\gamma_{01} - \gamma_{02} Y_{i0})}(1 - Y_{i1})$$
$$\phi_\theta^{1|1,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - (1 - Y_{i1}) - (1 - e^{\gamma_1 + \gamma_2 - (\gamma_1 Y_{i0} + \gamma_2 Y_{i-1})})\left(1 - Y_{i2}e^{\gamma_{02}(1-Y_{i3})(1-Y_{i0})}\right)Y_{i1}$$

Moreover, we have

$$\phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = (1 - Y_{i1})e^{Y_{i2}(\gamma_{01}Y_{i0} - \gamma_{02}(Y_{i0} - Y_{i-1})}$$
$$\phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = Y_{i1}e^{(1-Y_{i2})\left(\gamma_{01}(1-Y_{i0}) + \gamma_{02}(Y_{i0} - Y_{i-1})\right)}$$

give:

$$E\left[\phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0)|Y_{i-1}, Y_{i0}, A_i\right] = \pi^{0|0,Y_{i0}}(A_i) = \frac{1}{1 + e^{\gamma_{02}Y_{i0} + A_i}}$$
$$E\left[\phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0)|Y_{i-1}, Y_{i0}, A_i\right] = \pi^{1|1,Y_{i0}}(A_i) = \frac{e^{\gamma_{01} + \gamma_{02}Y_{i0} + A_i}}{1 + e^{\gamma_{01} + \gamma_{02}Y_{i0} + A_i}}$$

For $\pi^{0|0,0}(A_i)$ and $\pi^{0|0,Y_{i0}}(A_i)$ to match, we require $Y_{i0} = 0$ in which case:

$$\phi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} - (1 - e^{\gamma_2 Y_{i-1}})Y_{i2}(1 - Y_{i1})$$
$$\phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = (1 - Y_{i1})e^{\gamma_{02}Y_{i2}Y_{i-1}} = (1 - Y_{i1})Y_{i2}e^{\gamma_{02}Y_{i-1}} + (1 - Y_{i1})(1 - Y_{i2})$$

Therefore,

$$\psi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = \phi_\theta^{0|0,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) - \phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = 0$$

So there is no information about γ_1, γ_2 in this moment.

For $\pi^{0|0,1}(A_i)$ and $\pi^{0|0,Y_{i0}}(A_i)$ to match, we require $Y_{i0} = 1$ in which case:

$$\begin{aligned}
\phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) &= 1 - Y_{i1} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})(1 - Y_{i2})e^{\gamma_{01} Y_{i3}} Y_{i1} \\
\phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) &= (1 - Y_{i1})e^{Y_{i2}(\gamma_{01} - \gamma_{02}(1-Y_{i-1}))}
\end{aligned}$$

$$\phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - Y_{i1} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})(1 - Y_{i2})e^{\gamma_{01} Y_{i3}} Y_{i1}$$

Then, a valid moment condition that depends on γ_{01} and γ_{02} is

$$\begin{aligned}
&\psi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = \phi_\theta^{0|0,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) - \phi_\theta^{0|0}(Y_{i2}, Y_{i1}, Y_{i-1}^0) \\
&= 1 - Y_{i1} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})(1 - Y_{i2})e^{\gamma_{01} Y_{i3}} Y_{i1} - (1 - Y_{i1})e^{Y_{i2}(\gamma_{01} - \gamma_{02}(1-Y_{i-1}))} \\
&= 1 - Y_{i1} + e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})Y_{i3} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) \\
&- e^{\gamma_{01} - \gamma_{02}(1-Y_{i-1})}(1 - Y_{i1})Y_{i2} - (1 - Y_{i1})(1 - Y_{i2}) \\
&= e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})Y_{i3} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) \\
&- (e^{\gamma_{01} - \gamma_{02}(1-Y_{i-1})} - 1)(1 - Y_{i1})Y_{i2} \\
&= e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})Y_{i3} + (1 - e^{-\gamma_1 + \gamma_2(1-Y_{i-1})})Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) \\
&- e^{\gamma_{01} - \gamma_{02}(1-Y_{i-1})}(1 - e^{-\gamma_{01} + \gamma_{02}(1-Y_{i-1})})(1 - Y_{i1})Y_{i2}
\end{aligned}$$

It is useful to rescale this moment function by $\left(e^{\gamma_{01} - \gamma_{02}(1-Y_{i-1})}(1 - e^{-\gamma_{01} + \gamma_{02}(1-Y_{i-1})})\right)^{-1}$ in which case it simplifies to:

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = e^{\gamma_{02}(1-Y_{i-1})}Y_{i1}(1 - Y_{i2})Y_{i3} + e^{-\gamma_{01} + \gamma_{02}(1-Y_{i-1})}Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) - (1 - Y_{i1})Y_{i2}$$

Then, for the initial condition $Y_{i0} = 1, Y_{i-1} = 1$, we have

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, Y_{i1}, 1, 1) = Y_{i1}(1 - Y_{i2})Y_{i3} + e^{-\gamma_{01}}Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) - (1 - Y_{i1})Y_{i2}$$

Since, this moment function is strictly monotonic in γ_1 it identifies γ_1 uniquely. This moment coincides with m_{11} of Honoré and Weidner (2020). For the initial condition $Y_{i0} = 1, Y_{i-1} = 0$, we get

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = e^{\gamma_{02}}Y_{i1}(1 - Y_{i2})Y_{i3} + e^{-\gamma_{01} + \gamma_{02}}Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) - (1 - Y_{i1})Y_{i2}$$

This moment function is strictly monotonic in γ_2, thus provided that γ_1 is identified, it uniquely identifies

γ_2. This moment coincides with $m_{0,1}$ in Honoré and Weidner (2020).

Likewise, for $\pi^{1|1,0}(A_i)$ and $\pi^{0|0,Y_{i0}}(A_i)$ to match, we require $Y_{i0} = 0$ in which case:

$$\phi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = Y_{i1} + (1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}}) Y_{i2} e^{\gamma_{01}(1-Y_{i3})}(1 - Y_{i1})$$
$$\phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = Y_{i1} e^{(1-Y_{i2})\left(\gamma_1 - \gamma_2 Y_{i-1}\right)}$$

A valid moment condition that depends on γ_1, γ_2 is then,

$$\begin{aligned}
&\psi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = \phi_\theta^{1|1,0}(Y_{i3}, Y_{i2}, Y_{i-1}^1) - \phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0) \\
&= Y_{i1} + (1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}}) Y_{i2} e^{\gamma_{01}(1-Y_{i3})}(1 - Y_{i1}) - Y_{i1} e^{(1-Y_{i2})\left(\gamma_1 - \gamma_2 Y_{i-1}\right)} \\
&= Y_{i1} + e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + (1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2} Y_{i3} \\
&- Y_{i1}(1 - Y_{i2}) e^{\gamma_1 - \gamma_2 Y_{i-1}} - Y_{i1} Y_{i2} \\
&= e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + (1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2} Y_{i3} \\
&- (e^{\gamma_1 - \gamma_2 Y_{i-1}} - 1) Y_{i1}(1 - Y_{i2}) \\
&= e^{\gamma_{01}}(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + (1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}})(1 - Y_{i1}) Y_{i2} Y_{i3} \\
&- e^{\gamma_1 - \gamma_2 Y_{i-1}}(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1}}) Y_{i1}(1 - Y_{i2})
\end{aligned}$$

It is useful to rescale this moment function by $\left(e^{\gamma_{01} - \gamma_{02} Y_{i-1}}(1 - e^{-\gamma_{01} + \gamma_{02} Y_{i-1}})\right)^{-1}$ in which case it simplifies to:

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = e^{\gamma_{02} Y_{i-1}}(1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + e^{-\gamma_{01} + \gamma_{02} Y_{i-1}}(1 - Y_{i1}) Y_{i2} Y_{i3} - Y_{i1}(1 - Y_{i2})$$

Then, for the initial condition $Y_{i0} = 0, Y_{i-1} = 0$, we have

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i3}, Y_{i2}, Y_{i1}, 0, 0) = (1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + e^{-\gamma_{01}}(1 - Y_{i1}) Y_{i2} Y_{i3} - Y_{i1}(1 - Y_{i2})$$

Since, this moment function is strictly monotonic in γ_1 it identifies γ_1 uniquely. This moment coincides with m_{00} of Honoré and Weidner (2020). For the initial condition $Y_{i0} = 0, Y_{i-1} = 1$, we have

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i3}, Y_{i2}, Y_{i1}, 0, 1) = e^{\gamma_{02}}(1 - Y_{i1}) Y_{i2}(1 - Y_{i3}) + e^{-\gamma_{01} + \gamma_{02}}(1 - Y_{i1}) Y_{i2} Y_{i3} - Y_{i1}(1 - Y_{i2})$$

This moment function is strictly monotonic in γ_2, thus provided that γ_1 is identified, it uniquely identifies γ_2. This moment coincides with $m_{1,0}$ in Honoré and Weidner (2020).

Finally, for $\pi^{1|1,1}(A_i)$ and $\pi^{1|1,Y_{i0}}(A_i)$ to match, we require $Y_{i0} = 1$ in which case:

$$\phi_\theta^{1|1,1}(Y_{i3}, Y_{i2}, Y_{i-1}^1) = 1 - (1 - Y_{i1}) - (1 - e^{\gamma_2(1-Y_{i-1})})(1 - Y_{i2}) Y_{i1}$$
$$\phi_\theta^{1|1}(Y_{i2}, Y_{i1}, Y_{i-1}^0) = Y_{i1} e^{\gamma_{02}(1-Y_{i2})(1-Y_{i-1})}$$

but then the candidate moment function

$$
\begin{aligned}
&\psi_\theta^{1|1,1}(Y_{i3},Y_{i2},Y_{i-1}^1)=\phi_\theta^{1|1,1}(Y_{i3},Y_{i2},Y_{i-1}^1)-\phi_\theta^{1|1}(Y_{i2},Y_{i1},Y_{i-1}^0)\\
&=1-(1-Y_{i1})-(1-e^{\gamma_2(1-Y_{i-1})})(1-Y_{i2})Y_{i1}-Y_{i1}e^{\gamma_{02}(1-Y_{i2})(1-Y_{i-1})}\\
&=Y_{i1}-(1-e^{\gamma_2(1-Y_{i-1})})(1-Y_{i2})Y_{i1}-Y_{i1}(1-Y_{i2})e^{\gamma_2(1-Y_{i-1})}-Y_{i1}Y_{i2}\\
&=0
\end{aligned}
$$

so there is no identifying information in this moment function for the parameters γ_1,γ_2.

I Identification of the AR(2) with strictly exogenous regressors

I.1 Identification for $T=3$ with variability in the initial condition

By Theorem 1, the transition functions associated to: $\pi_2^{0|0,0}(A_i,X_i),\pi_2^{0|0,1}(A_i,X_i),\pi_2^{1|1,0}(A_i,X_i),\pi_2^{1|1,1}(A_i,X_i)$ are given by:

$$
\begin{aligned}
&\phi_\theta^{0|0}(Y_{i3},Y_{i2},Y_{i0}^1,X_i)=(1-Y_{i2})e^{Y_{i3}(\gamma_1Y_{i1}-\gamma_2(Y_{i1}-Y_{i0})-X_{i32}'\beta)}\\
&\phi_\theta^{1|1}(Y_{i3},Y_{i2},Y_0^1,X_i)=Y_{i2}e^{(1-Y_{i3})\left(\gamma_1(1-Y_{i1})+\gamma_2(Y_{i1}-Y_{i0})+X_{i32}'\beta\right)}\\
&\phi_\theta^{0|0,0}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=1-Y_{i1}-\left(1-e^{\gamma_1Y_{i0}+\gamma_2Y_{i-1}-X_{i31}'\beta}\right)\left(1-\phi_\theta^{0|0}(Y_{i3},Y_{i2},Y_{i0}^1,X_i)\right)(1-Y_{i1})\\
&\phi_\theta^{0|0,1}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=(1-Y_{i1})+\left(1-e^{-\gamma_1Y_{i0}+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)\phi_\theta^{0|0}(Y_{i3},Y_{i2},Y_{i0}^1,X_i)Y_{i1}\\
&\phi_\theta^{1|1,1}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=Y_{i1}-\left(1-e^{\gamma_1(1-Y_{i0})+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)\left(1-\phi_\theta^{1|1}(Y_{i3},Y_{i2},Y_0^1,X_i)\right)Y_{i1}\\
&\phi_\theta^{1|1,0}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=Y_{i1}+\left(1-e^{-\gamma_1(1-Y_{i0})+\gamma_2Y_{i-1}-X_{i31}'\beta}\right)\phi_\theta^{1|1}(Y_{i3},Y_{i2},Y_0^1,X_i)(1-Y_{i1})
\end{aligned}
$$

After simplification

$$
\begin{aligned}
&\phi_\theta^{0|0,0}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=e^{\gamma_1Y_{i0}+\gamma_2Y_{i-1}-X_{i31}'\beta}(1-Y_{i1})+\left(1-e^{\gamma_1Y_{i0}+\gamma_2Y_{i-1}-X_{i31}'\beta}\right)(1-Y_{i1})(1-Y_{i2})e^{Y_{i3}(\gamma_2Y_{i0}-X_{i32}'\beta)}\\
&\phi_\theta^{0|0,1}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=(1-Y_{i1})+\left(1-e^{-\gamma_1Y_{i0}+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)Y_{i1}(1-Y_{i2})e^{Y_{i3}(\gamma_1-\gamma_2(1-Y_{i0})-X_{i32}'\beta)}\\
&\phi_\theta^{1|1,1}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=e^{\gamma_1(1-Y_{i0})+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}Y_{i1}+\left(1-e^{\gamma_1(1-Y_{i0})+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)Y_{i1}Y_{i2}e^{(1-Y_{i3})(\gamma_2(1-Y_{i0})+X_{i32}'\beta)}\\
&\phi_\theta^{1|1,0}(Y_{i3},Y_{i2},Y_{i-1}^1,X_i)=Y_{i1}+\left(1-e^{-\gamma_1(1-Y_{i0})+\gamma_2Y_{i-1}-X_{i31}'\beta}\right)(1-Y_{i1})Y_{i2}e^{(1-Y_{i3})(\gamma_1-\gamma_2Y_{i0}+X_{i32}'\beta)}
\end{aligned}
$$

Moreover, we have

$$
\begin{aligned}
&\phi_\theta^{0|0}(Y_{i2},Y_{i1},Y_{i-1}^0,X_i)=(1-Y_{i1})e^{Y_{i2}(\gamma_1Y_{i0}-\gamma_2(Y_{i0}-Y_{i-1})-X_{i21}'\beta)}\\
&\phi_\theta^{1|1}(Y_{i2},Y_{i1},Y_{i-1}^0,X_i)=Y_{i1}e^{(1-Y_{i2})\left(\gamma_1(1-Y_{i0})+\gamma_2(Y_{i0}-Y_{i-1})+X_{i21}'\beta\right)}
\end{aligned}
$$

give:

$$E\left[\phi_\theta^{0|0}(Y_{i2},Y_{i1},Y_{i-1}^0,X_i)|Y_{i-1},Y_{i0},A_i\right] = \pi_1^{0|0,Y_{i0}}(A_i,X_i) = \frac{1}{1+e^{\gamma_2 Y_{i0}+X_{i2}'\beta+A_i}}$$
$$E\left[\phi_\theta^{1|1}(Y_{i2},Y_{i1},Y_{i-1}^0,X_i)|Y_{i-1},Y_{i0},A_i\right] = \pi_1^{1|1,Y_{i0}}(A_i,X_i) = \frac{e^{\gamma_1+\gamma_2 Y_{i0}+X_{i2}'\beta+A_i}}{1+e^{\gamma_1+\gamma_2 Y_{i0}++X_{i2}'\beta+A_i}}$$

For $\pi_2^{0|0,0}(A_i,X_i)$ and $\pi_1^{0|0,Y_{i0}}(A_i,X_i)$ to match, we require both $Y_{i0}=0$ and $X_{i3}=X_{i2}$ in which case:

$$\phi_\theta^{0|0,0}(Y_{i1}^3,0,Y_{i-1},X_i) = e^{\gamma_2 Y_{i-1}-X_{i31}'\beta}(1-Y_{i1}) + \left(1-e^{\gamma_2 Y_{i-1}-X_{i31}'\beta}\right)(1-Y_{i1})(1-Y_{i2})$$
$$\begin{aligned}\phi_\theta^{0|0}(Y_{i1}^2,0,Y_{i-1},X_i) &= (1-Y_{i1})e^{Y_{i2}(\gamma_2 Y_{i-1}-X_{i31}'\beta)}\\ &= (1-Y_{i1})Y_{i2}e^{\gamma_2 Y_{i-1}-X_{i31}'\beta} + (1-Y_{i1})(1-Y_{i2})\end{aligned}$$

Therefore,

$$\psi_\theta^{0|0,0}(Y_{i1}^3,0,Y_{i-1},X_i) = \phi_\theta^{0|0,0}(Y_{i1}^3,0,Y_{i-1},X_i) - \phi_\theta^{0|0}(Y_{i1}^2,0,Y_{i-1},X_i) = 0$$

So there is no information about the model parameters in this moment function.

For $\pi_2^{0|0,1}(A_i,X_i)$ and $\pi_1^{0|0,Y_{i0}}(A_i,X_i)$ to match, we require both $Y_{i0}=1$ and $X_{i3}=X_{i2}$ in which case:

$$\phi_\theta^{0|0,1}(Y_{i1}^3,1,Y_{i-1},X_i) = (1-Y_{i1}) + \left(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)Y_{i1}(1-Y_{i2})e^{\gamma_1 Y_{i3}}$$
$$\phi_\theta^{0|0}(Y_{i1}^2,1,Y_{i-1},X_i) = (1-Y_{i1})e^{Y_{i2}(\gamma_1-\gamma_2(1-Y_{i-1})-X_{i31}'\beta)}$$

Then, a valid moment condition that depends on all model parameters is:

$$\begin{aligned}
&\psi_\theta^{0|0,1}(Y_{i1}^3,1,Y_{i-1},X_i) = \phi_\theta^{0|0,1}(Y_{i1}^3,1,Y_{i-1},X_i) - \phi_\theta^{0|0}(Y_{i1}^2,1,Y_{i-1},X_i)\\
&= (1-Y_{i1})\\
&+ \left(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)e^{\gamma_1}Y_{i1}(1-Y_{i2})Y_{i3}\\
&+ \left(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)Y_{i1}(1-Y_{i2})(1-Y_{i3})\\
&- e^{\gamma_1-\gamma_2(1-Y_{i-1})-X_{i31}'\beta}(1-Y_{i1})Y_{i2}\\
&- (1-Y_{i1})(1-Y_{i2})\\
&= \left(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)e^{\gamma_1}Y_{i1}(1-Y_{i2})Y_{i3}\\
&+ \left(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}\right)Y_{i1}(1-Y_{i2})(1-Y_{i3})\\
&- e^{\gamma_1-\gamma_2(1-Y_{i-1})-X_{i31}'\beta}(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta})(1-Y_{i1})Y_{i2}
\end{aligned}$$

Rescaling this moment function by the factor $\left(e^{\gamma_1-\gamma_2(1-Y_{i-1})-X_{i31}'\beta}(1-e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta})\right)^{-1}$, one obtains

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i1}^3,1,Y_{i-1},X_i) = e^{\gamma_2(1-Y_{i-1})+X_{i31}'\beta}Y_{i1}(1-Y_{i2})Y_{i3} + e^{-\gamma_1+\gamma_2(1-Y_{i-1})+X_{i31}'\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3}) - (1-Y_{i1})Y_{i2}$$

Thus, for for the initial condition $Y_{i0} = 1, Y_{i-1} = 1$, we have

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i1}^3, 1, 1, X_i) = e^{X'_{i31}\beta} Y_{i1}(1 - Y_{i2})Y_{i3} + e^{-\gamma_1 + X'_{i31}\beta} Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) - (1 - Y_{i1})Y_{i2}$$

which only depends on γ_1 and β. Clearly, it is strictly decreasing in γ_1 and by conditioning on particular sets for the regressors, it can be shown that it uniquely identifies γ_1, β (See Honoré and Weidner (2020)). Instead, for the initial condition $Y_{i0} = 1, Y_{i-1} = 0$, we have

$$\widetilde{\psi_\theta^{0|0,1}}(Y_{i1}^3, 1, 0, X_i) = e^{\gamma_2 + X'_{i31}\beta} Y_{i1}(1 - Y_{i2})Y_{i3} + e^{-\gamma_1 + \gamma_2 + X'_{i31}\beta} Y_{i1}(1 - Y_{i2})(1 - Y_{i3}) - (1 - Y_{i1})Y_{i2}$$

Provided that γ_1, β are identified, the strict monotonicity of the moment functions in γ_2 ensure that γ_2 is identified.

Analogously, for $\pi_2^{1|1,0}(A_i, X_i)$ and $\pi_1^{0|0,Y_{i0}}(A_i)$ to match, we require both $Y_{i0} = 0$ and $X_{i3} = X_{i2}$ in which case:

$$\begin{aligned}
\phi_\theta^{1|1,0}(Y_{i1}^3, 0, Y_{i-1}, X_i) &= Y_{i1} + \left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right)(1 - Y_{i1})Y_{i2} e^{\gamma_1(1 - Y_{i3})} \\
\phi_\theta^{1|1}(Y_{i1}^2, 0, Y_{i-1}, X_i) &= Y_{i1} e^{(1 - Y_{i2})\left(\gamma_1 - \gamma_2 Y_{i-1} + X'_{i31}\beta\right)}
\end{aligned}$$

Then, a valid moment function that depends on all model parameters is:

$$\begin{aligned}
\psi_\theta^{1|1,0}(Y_{i1}^3, 0, Y_{i-1}, X_i) &= \phi_\theta^{1|1,0}(Y_{i1}^3, 0, Y_{i-1}, X_i) - \phi_\theta^{1|1}(Y_{i1}^2, 0, Y_{i-1}, X_i) \\
&= Y_{i1} + \\
&+ \left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right) e^{\gamma_1}(1 - Y_{i1})Y_{i2}(1 - Y_{i3}) \\
&+ \left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right)(1 - Y_{i1})Y_{i2}Y_{i3} \\
&- e^{\gamma_1 - \gamma_2 Y_{i-1} + X'_{i31}\beta} Y_{i1}(1 - Y_{i2}) \\
&- Y_{i1}Y_{i2} \\
&= \left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right) e^{\gamma_1}(1 - Y_{i1})Y_{i2}(1 - Y_{i3}) \\
&+ \left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right)(1 - Y_{i1})Y_{i2}Y_{i3} \\
&- e^{\gamma_1 - \gamma_2 Y_{i-1} + X'_{i31}\beta}\left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right) Y_{i1}(1 - Y_{i2})
\end{aligned}$$

Rescaling this moment function by the factor $\left(e^{\gamma_1 - \gamma_2 Y_{i-1} + X'_{i31}\beta}\left(1 - e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}\right)\right)^{-1}$, one obtains

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, Y_{i-1}, X_i) = e^{\gamma_2 Y_{i-1} - X'_{i31}\beta}(1 - Y_{i1})Y_{i2}(1 - Y_{i3}) + e^{-\gamma_1 + \gamma_2 Y_{i-1} - X'_{i31}\beta}(1 - Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1 - Y_{i2})$$

For the initial condition $Y_{i0} = 0, Y_{i-1} = 0$, we have

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, 0, X_i) = e^{-X'_{i31}\beta}(1 - Y_{i1})Y_{i2}(1 - Y_{i3}) + e^{-\gamma_1 - X'_{i31}\beta}(1 - Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1 - Y_{i2})$$

This moment function also only depends on γ_1, β. Moreover, it is strictly decreasing in γ_1 and by using different conditioning sets for the regressors, one can show that it uniquely identifies γ_1, β. Instead, for the initial condition $Y_{i0} = 0, Y_{i-1} = 1$, we obtain

$$\widetilde{\psi_\theta^{1|1,0}}(Y_{i1}^3, 0, 1, X_i) = e^{\gamma_2 - X_{i31}'\beta}(1 - Y_{i1})Y_{i2}(1 - Y_{i3}) + e^{-\gamma_1 + \gamma_2 - X_{i31}'\beta}(1 - Y_{i1})Y_{i2}Y_{i3} - Y_{i1}(1 - Y_{i2})$$

Provided that γ_1, β is identified, the strict monotonicity of this moment function in γ_2 implies that it identifies γ_2 uniquely.

Lastly, for $\pi_2^{1|1,1}(A_i)$ and $\pi_1^{1|1,Y_{i0}}(A_i)$ to match, we require both $Y_{i0} = 1$ and $X_{i3} = X_{i2}$ in which case:

$$\begin{aligned}
\phi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1}, X_i) &= e^{\gamma_2(1-Y_{i-1}) + X_{i31}'\beta}Y_{i1} + \left(1 - e^{\gamma_2(1-Y_{i-1}) + X_{i31}'\beta}\right)Y_{i1}Y_{i2} \\
\phi_\theta^{1|1}(Y_{i1}^2, 1, Y_{i-1}, X_i) &= Y_{i1}e^{(1-Y_{i2})\left(\gamma_2(1-Y_{i-1}) + X_{i21}'\beta\right)} \\
&= Y_{i1}(1 - Y_{i2})e^{\gamma_2(1-Y_{i-1}) + X_{i21}'\beta} + Y_{i1}Y_{i2}
\end{aligned}$$

Then, a valid moment function

$$\begin{aligned}
\psi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1}, X_i) &= \phi_\theta^{1|1,1}(Y_{i1}^3, 1, Y_{i-1}, X_i) - \phi_\theta^{1|1}(Y_{i1}^2, 1, Y_{i-1}, X_i) \\
&= 0
\end{aligned}$$

is identically zero and hence contains no information about the model parameters.

I.2 Identification at infinity with $T = 4$

We recall from the discussion of Section 4.8 that $T = 4$ and $K_x \geq 2$ so that there are at least 2 exogenous explanatory variables. We have $X_{it} = (W_{it}, R_{it}')' \in \mathbb{R}^{K_x}$, $\beta = (\beta_W, \beta_R')' \in \mathbb{R}^{K_x}$ and $Z_i = (R_i', W_{i1}, W_{i3}, W_{i4})' \in \mathbb{R}^{4K_x - 1}$. Our goal is to prove Theorem 3 under Assumptions 1 and 2.

Preliminary results. By Theorem 1, the transition functions associated to: $\pi_3^{0|0,0}(A_i, X_i), \pi_3^{0|0,1}(A_i, X_i), \pi_3^{1|1,0}(A_i, X_i)$,
$\pi_3^{1|1,1}(A_i, X_i)$ are given by:

$$\begin{aligned}
\phi_\theta^{0|0}(Y_{i4}, Y_{i3}, Y_{i1}^2, X_i) &= (1 - Y_{i3})e^{Y_{i4}(\gamma_1 Y_{i2} - \gamma_2(Y_{i2} - Y_{i1}) - X_{i43}'\beta)} \\
\phi_\theta^{1|1}(Y_{i4}, Y_{i3}, Y_{i1}^2, X_i) &= Y_{i3}e^{(1-Y_{i4})\left(\gamma_1(1-Y_{i2}) + \gamma_2(Y_{i2} - Y_{i1}) + X_{i43}'\beta\right)} \\
\phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) &= 1 - Y_{i2} - \left(1 - e^{\gamma_1 Y_{i1} + \gamma_2 Y_{i0} - X_{i42}'\beta}\right)\left(1 - \phi_\theta^{0|0}(Y_{i4}, Y_{i3}, Y_{i1}^2, X_i)\right)(1 - Y_{i2}) \\
\phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) &= (1 - Y_{i2}) + \left(1 - e^{-\gamma_1 Y_{i1} + \gamma_2(1-Y_{i0}) + X_{i42}'\beta}\right)\phi_\theta^{0|0}(Y_{i4}, Y_{i3}, Y_{i1}^2, X_i)Y_{i2} \\
\phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) &= Y_{i2} - \left(1 - e^{\gamma_1(1-Y_{i1}) + \gamma_2(1-Y_{i0}) + X_{i42}'\beta}\right)\left(1 - \phi_\theta^{1|1}(Y_{i4}, Y_{i3}, Y_1^2, X_i)\right)Y_{i2} \\
\phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) &= Y_{i2} + \left(1 - e^{-\gamma_1(1-Y_{i1}) + \gamma_2 Y_{i0} - X_{i42}'\beta}\right)\phi_\theta^{1|1}(Y_{i4}, Y_{i3}, Y_1^2, X_i)(1 - Y_{i2})
\end{aligned}$$

After simplification, one gets

$$\phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) = e^{\gamma_1 Y_{i1} + \gamma_2 Y_{i0} - X_{i42}'\beta}(1 - Y_{i2}) + \left(1 - e^{\gamma_1 Y_{i1} + \gamma_2 Y_{i0} - X_{i42}'\beta}\right)(1 - Y_{i2})(1 - Y_{i3})e^{Y_{i4}(\gamma_2 Y_{i1} - X_{i43}'\beta)}$$
$$\phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) = (1 - Y_{i2}) + \left(1 - e^{-\gamma_1 Y_{i1} + \gamma_2(1 - Y_{i0}) + X_{i42}'\beta}\right) Y_{i2}(1 - Y_{i3})e^{Y_{i4}(\gamma_1 - \gamma_2(1 - Y_{i1}) - X_{i43}'\beta)}$$
$$\phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) = e^{\gamma_1(1 - Y_{i1}) + \gamma_2(1 - Y_{i0}) + X_{i42}'\beta} Y_{i2} + \left(1 - e^{\gamma_1(1 - Y_{i1}) + \gamma_2(1 - Y_{i0}) + X_{i42}'\beta}\right) Y_{i2} Y_{i3} e^{(1 - Y_{i4})\left(\gamma_2(1 - Y_{i1}) + X_{i43}'\beta\right)}$$
$$\phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) = Y_{i2} + \left(1 - e^{-\gamma_1(1 - Y_{i1}) + \gamma_2 Y_{i0} - X_{i42}'\beta}\right)(1 - Y_{i2}) Y_{i3} e^{(1 - Y_{i4})(\gamma_1 - \gamma_2 Y_{i1} + X_{i43}'\beta)}$$

Then, Lemma 4 and Proposition 3 imply that a set of valid moment functions for arbitrary regressors are:

$$\begin{aligned}
\psi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) &= \phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - \zeta_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) \\
&= \phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - (1 - Y_{i1}) - \left(1 - e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X_{i41}'\beta}\right)\phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) Y_{i1} \\
&= (1 - Y_{i1})\phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) + e^{-\gamma_1 Y_{i0} - \gamma_2 Y_{i-1} + X_{i41}'\beta} Y_{i1}\phi_\theta^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - (1 - Y_{i1})
\end{aligned}$$

$$\begin{aligned}
\psi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) &= \phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - \zeta_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) \\
&= \phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - (1 - Y_{i1}) - \left(1 - e^{-\gamma_1 Y_{i0} + \gamma_2(1 - Y_{i-1}) + X_{i41}'\beta}\right)\phi_\theta^{0|01}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) Y_{i1} \\
&= (1 - Y_{i1})\phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) + e^{-\gamma_1 Y_{i0} + \gamma_2(1 - Y_{i-1}) + X_{i41}'\beta}\phi_\theta^{0|0,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - (1 - Y_{i1})
\end{aligned}$$

$$\begin{aligned}
\psi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) &= \phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - \zeta_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) \\
&= \phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - Y_{i1} - \left(1 - e^{-\gamma_1(1 - Y_{i0}) - \gamma_2(1 - Y_{i-1}) - X_{i41}'\beta}\right)\phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i)(1 - Y_{i1}) \\
&= Y_{i1}\phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) + e^{-\gamma_1(1 - Y_{i0}) - \gamma_2(1 - Y_{i-1}) - X_{i41}'\beta}(1 - Y_{i1})\phi_\theta^{1|1,1}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - Y_{i1}
\end{aligned}$$

$$\begin{aligned}
\psi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) &= \phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - \zeta_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, X_i) \\
&= \phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - Y_{i1} - \left(1 - e^{-\gamma_1(1 - Y_{i0}) + \gamma_2 Y_{i-1} - X_{i41}'\beta}\right)\phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i)(1 - Y_{i1}) \\
&= Y_{i1}\phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) + e^{-\gamma_1(1 - Y_{i0}) + \gamma_2 Y_{i-1} - X_{i41}'\beta}(1 - Y_{i1})\phi_\theta^{1|1,0}(Y_{i4}, Y_{i3}, Y_{i0}^2, X_i) - Y_{i1}
\end{aligned}$$

which after tedious simplifications and rearrangement yield

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

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

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

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

Proof of Theorem 3. Define, the "limiting" moment function

$$
\begin{aligned}
\psi_{\theta,\infty}^{0|0,0}(Y_{i4}, Y_{i3}, Y_{i-1}^2, Z_i) = &-(1-Y_{i1})(1-Y_{i2})Y_{i3} \\
&+ \left[e^{X'_{i34}\beta} - 1\right](1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4} \\
&+ e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})+X'_{i31}\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4} \\
&+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}+X'_{i41}\beta} Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})
\end{aligned}
\tag{17}
$$

For $s \in \{-,+\}^{K_x}$, consider the moment objective

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

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

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

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

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

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

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

Hence another application of the Dominated Convergence Theorem gives

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

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, $w_2\mapsto E\left[\psi_{\theta,-\infty}^{0|0,0}(Y_{i4},Y_{i3},Y_{i-1}^2,Z_i)|Y_i^0=y^0,X_i\in\mathcal{X}_s,W_{i2}=w_2\right]$ 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 (10) 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

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

Similarly,

$$
\begin{aligned}
&\frac{\partial \Psi^{0|0,0}_{s,0,0}(\theta)}{\partial \beta_k} \\
&= E\left[\frac{\partial \psi^{0|0,0}_{\theta,-\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i)}{\partial \beta_k} | Y^0_i = (0,0), X_i \in \mathcal{X}_s, W_{i2} = \infty\right] \\
&= E\left[X_{ik,34} e^{X'_{i34}\beta} \underbrace{E\left[(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4} | Y^0_i = (0,0), Z_i, W_{i2} = \infty, A_i\right]}_{>0} | Y^0_i = (0,0), X_i \in \mathcal{X}_s, W_{i2} = \infty\right] \\
&+ E\left[X_{ik,31} e^{\gamma_2 + X'_{i31}\beta} \underbrace{E\left[Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4} | Y^0_i = (0,0), Z_i, W_{i2} = \infty, A_i\right]}_{>0} | Y^0_i = (0,0), X_i \in \mathcal{X}_s, W_{i2} = \infty\right] \\
&+ E\Bigg[X_{ik,41} e^{\gamma_2 + X'_{i31}\beta} \times \\
&\underbrace{E\left[Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4}) | Y^0_i = (0,0), Z_i \in \mathcal{X}_s, W_{i2} = \infty, A_i\right]}_{>0} | Y^0_i = (0,0), X_i \in \mathcal{X}_s, W_{i2} = \infty\Bigg]
\end{aligned}
$$

The last display shows that $\frac{\partial \Psi^{0|0,0}_{s,0,0}(\theta)}{\partial \beta_k} > 0$ if $s_k = +$ and $\frac{\partial \Psi^{0|0,0}_{s,0,0}(\theta)}{\partial \beta_k} < 0$ 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:

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

has at most one solution. It is precisely (γ_{02}, β_0), since the validity of $\psi^{0|0,0}_{\theta}(Y_{i4}, Y_{i3}, Y^2_{i-1}, X_i)$ for arbitrary X_i directly implies the validity of the limiting moment $\psi^{0|0,0}_{\theta,-\infty}(Y_{i4}, Y_{i3}, Y^2_{i-1}, Z_i)$ at "$W_{i2} = \infty$". Then, notice that for any other initial condition $y^0 \in \{(0,1),(1,0),(1,1)\}$, the objective $\Psi^{0|0,0}_{s,y^0}(\theta)$ is strictly monotonic in γ_1. Hence, given (γ_{02}, β_0), it point identifies γ_{01}. This concludes the proof of Theorem 3.

J Proof of Lemma 5

The proposed functional form for the transition function $\phi_\theta^{(0,0)|(0,0)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq (0,0)$. Hence

$$
\begin{aligned}
&E\left[\phi_\theta^{(0,0)|(0,0)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \\
&\frac{1}{(1+e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X'_{1,it}\beta_1+A_{1,i}})}\frac{1}{(1+e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X'_{2,it}\beta_2+A_{2,i}})}\times \\
&\left(\frac{e^{X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{X'_{1,it+1}\beta_1+A_{1,i}}}\frac{e^{X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,0)|(0,0)}((1,1),(0,0),Y_{it-1},X_i)+\right. \\
&\frac{e^{X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{X'_{1,it+1}\beta_1+A_{1,i}}}\frac{1}{1+e^{X'_{2,it+1}\beta_2+A_{1,i}}}\phi_\theta^{(0,0)|(0,0)}((1,0),(0,0),Y_{it-1},X_i)+ \\
&\frac{1}{1+e^{X'_{1,it+1}\beta_1+A_{1,i}}}\frac{e^{X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,0)|(0,0)}((0,1),(0,0),Y_{it-1},X_i)+ \\
&\left.\frac{1}{1+e^{X'_{1,it+1}\beta_1+A_{1,i}}}\frac{1}{1+e^{X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,0)|(0,0)}((0,0),(0,0),Y_{it-1},X_i)\right)
\end{aligned}
$$

Thus, to obtain the transition probability $\pi_t^{(0,0|(0,0)}(A_i, X_i) = \frac{1}{1+e^{X_{1,it+1}\beta_1+A_{1,i}}}\frac{1}{1+e^{X_{2,it+1}\beta_2+A_{2,i}}}$, we must set:

$$
\begin{aligned}
&\phi_\theta^{(0,0)|(0,0)}((1,1),(0,0),Y_{it-1},X_i) = e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+(X_{1,it}-X_{1,it+1})\beta_1+\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+(X_{2,it}-X_{2,it+1})\beta_2} \\
&\phi_\theta^{(0,0)|(0,0)}((1,0),(0,0),Y_{it-1},X_i) = e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+(X_{1,it}-X_{1,it+1})\beta_1} \\
&\phi_\theta^{(0,0)|(0,0)}((0,1),(0,0),Y_{it-1},X_i) = e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+(X_{2,it}-X_{2,it+1})\beta_2} \\
&\phi_\theta^{(0,0)|(0,0)}((0,0),(0,0),Y_{it-1},X_i) = 1 \\
&\qquad \phi_\theta^{(0,0)|(0,0)}(k,l,Y_{it-1},X_i) = 0, \quad \forall k \in \mathcal{Y}, \quad \forall l \in \mathcal{Y}\setminus\{(0,0)\}
\end{aligned}
$$

This can be expressed compactly as:

$$
\begin{aligned}
&\phi_\theta^{(0,0)|(0,0)}(Y_{it-1}^{t+1}, X_i) = (1-Y_{1,it})(1-Y_{2,it})\times \\
&e^{Y_{1,it+1}\left(\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}-\Delta X'_{1,it+1}\beta_1\right)+Y_{2,it+1}\left(\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}-\Delta X'_{2,it+1}\beta_2\right)}
\end{aligned}
\tag{18}
$$

Similarly, the proposed functional form for the transition function $\phi_\theta^{(1,0)|(1,0)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq (1,0)$. Hence

$$\begin{aligned}
&E\left[\phi_\theta^{(1,0)|(1,0)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \\
&\frac{e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X_{1,it}\beta_1+A_{1,i}}}{(1+e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X_{1,it}\beta_1+A_{1,i}})} \frac{1}{(1+e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X_{2,it}\beta_2+A_{2,i}})} \times \\
&\Bigg(\frac{e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}}{1+e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}} \frac{e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}}{1+e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}} \phi_\theta^{(1,0)|(1,0)}((1,1),(1,0),Y_{it-1},X_i) + \\
&\frac{e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}}{1+e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}} \frac{1}{1+e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{1,i}}} \phi_\theta^{(1,0)|(1,0)}((1,0),(1,0),Y_{it-1},X_i) + \\
&\frac{1}{1+e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}} \frac{e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}}{1+e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}} \phi_\theta^{(1,0)|(1,0)}((0,1),(1,0),Y_{it-1},X_i) + \\
&\frac{1}{1+e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}} \frac{1}{1+e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}} \phi_\theta^{(1,0)|(1,0)}((0,0),(1,0),Y_{it-1},X_i)\Bigg)
\end{aligned}$$

Thus, to obtain the transition probability $\pi_t^{(1,0)|(1,0)}(A_i, X_i) = \frac{e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}}{1+e^{\gamma_{11}+X_{1,it+1}'\beta_1+A_{1,i}}} \frac{1}{1+e^{\gamma_{21}+X_{2,it+1}'\beta_2+A_{2,i}}}$, we must set:

$$\begin{aligned}
&\phi_\theta^{(1,0)|(1,0)}((1,1),(1,0),Y_{it-1},X_i) = e^{\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}Y_{2,it-1}-\Delta X_{2,it+1}\beta_2} \\
&\phi_\theta^{(1,0)|(1,0)}((1,0),(1,0),Y_{it-1},X_i) = 1 \\
&\phi_\theta^{(1,0)|(1,0)}((0,1),(1,0),Y_{it-1},X_i) = e^{-\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}Y_{2,it-1}-\Delta X_{1,it}\beta_1\right)+\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}Y_{2,it-1}-\Delta X_{2,it+1}\beta_2} \\
&\phi_\theta^{(1,0)|(1,0)}((0,0),(1,0),Y_{it-1},X_i) = e^{-\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}Y_{2,it-1}-\Delta X_{1,it+1}\beta_1\right)} \\
&\qquad \phi_\theta^{(1,0)|(1,0)}(k,l,Y_{it-1},X_i) = 0, \quad \forall k \in \mathcal{Y}, \quad \forall l \in \mathcal{Y} \setminus \{(1,0)\}
\end{aligned}$$

This can be expressed compactly as:

$$\begin{aligned}
&\phi_\theta^{(1,0)|(1,0)}(Y_{it-1}^{t+1}, X_i) = Y_{1,it}(1-Y_{2,it}) \\
&e^{-(1-Y_{1,it+1})\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}Y_{2,it-1}-\Delta X_{1,it}\beta_1\right)+Y_{2,it+1}\left(\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}Y_{2,it-1}-\Delta X_{2,it+1}\beta_2\right)}
\end{aligned} \tag{19}$$

Analogously, the proposed functional form for the transition function $\phi_\theta^{(0,1)|(0,1)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq (0,1)$. Hence

$$
\begin{aligned}
&E\left[\phi_\theta^{(0,1)|(0,1)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \\
&\frac{1}{(1+e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X_{1,it}\beta_1+A_{1,i}})}\frac{e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X_{2,it}\beta_2+A_{2,i}}}{(1+e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X_{2,it}\beta_2+A_{2,i}})}\times \\
&\left(\frac{e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}\frac{e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,1)|(0,1)}((1,1),(0,1),Y_{it-1},X_i)+\right. \\
&\frac{e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}\frac{1}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{1,i}}}\phi_\theta^{(0,1)|(0,1)}((1,0),(0,1),Y_{it-1},X_i)+ \\
&\frac{1}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}\frac{e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,1)|(0,1)}((0,1),(0,1),Y_{it-1},X_i)+ \\
&\left.\frac{1}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}\frac{1}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}\phi_\theta^{(0,1)|(0,1)}((0,0),(0,1),Y_{it-1},X_i)\right)
\end{aligned}
$$

Thus, to obtain the transition probability $\pi_t^{(0,1)|(0,1)}(A_i, X_i) = \frac{1}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}\frac{e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}$, we must set:

$$
\begin{aligned}
\phi_\theta^{(0,1)|(0,1)}((1,1),(0,1),Y_{it-1},X_i) &= e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1} \\
\phi_\theta^{(0,1)|(0,1)}((1,0),(0,1),Y_{it-1},X_i) &= e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1-\left(\gamma_{21}Y_{1,it-1}+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)} \\
\phi_\theta^{(0,1)|(0,1)}((0,1),(0,1),Y_{it-1},X_i) &= 1 \\
\phi_\theta^{(0,1)|(0,1)}((0,0),(0,1),Y_{it-1},X_i) &= e^{-\left(\gamma_{21}Y_{1,it-1}+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)} \\
\phi_\theta^{(0,1)|(0,1)}(k,l,Y_{it-1},X_i) &= 0, \quad \forall k \in \mathcal{Y}, \quad \forall l \in \mathcal{Y} \setminus \{(0,1)\}
\end{aligned}
$$

This can be expressed compactly as:

$$
\begin{aligned}
&\phi_\theta^{(0,1)|(0,1)}(Y_{it-1}^{t+1}, X_i) = (1-Y_{1,it})Y_{2,it}\times \\
&e^{Y_{1,it+1}\left(\gamma_{11}Y_{1,it-1}+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1\right)-(1-Y_{2,it+1})\left(\gamma_{21}Y_{1,it-1}+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)}
\end{aligned}
\tag{20}
$$

Finally, the proposed functional form for the transition function $\phi_\theta^{(1,1)|(1,1)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq (1,1)$. Hence

$$\begin{aligned}
&E\left[\phi_\theta^{(1,1)|(1,1)}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i, A_i\right] = \\
&\frac{e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X_{1,it}\beta_1+A_{1,i}}}{(1+e^{\gamma_{11}Y_{1,it-1}+\gamma_{12}Y_{2,it-1}+X_{1,it}\beta_1+A_{1,i}})} \frac{e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X_{2,it}\beta_2+A_{2,i}}}{(1+e^{\gamma_{21}Y_{1,it-1}+\gamma_{22}Y_{2,it-1}+X_{2,it}\beta_2+A_{2,i}})} \times \\
&\Bigg(\frac{e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}} \frac{e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}} \phi_\theta^{(1,1)|(1,1)}((1,1),(1,1),Y_{it-1},X_i) + \\
&\frac{e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}} \frac{1}{1+e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{1,i}}} \phi_\theta^{(1,1)|(1,1)}((1,0),(1,1),Y_{it-1},X_i) + \\
&\frac{1}{1+e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}} \frac{e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}} \phi_\theta^{(1,1)|(1,1)}((0,1),(1,1),Y_{it-1},X_i) + \\
&\frac{1}{1+e^{\gamma_{11}+\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}} \frac{1}{1+e^{\gamma_{21}+\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}} \phi_\theta^{(1,1)|(1,1)}((0,0),(1,1),Y_{it-1},X_i) \Bigg)
\end{aligned}$$

Thus, to obtain the transition probability $\pi_t^{(1,1)|(1,1)}(A_i, X_i) = \frac{e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}}{1+e^{\gamma_{12}+X'_{1,it+1}\beta_1+A_{1,i}}} \frac{e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}{1+e^{\gamma_{22}+X'_{2,it+1}\beta_2+A_{2,i}}}$, we must set:

$$\begin{aligned}
&\phi_\theta^{(1,1)|(1,1)}((1,1),(1,1),Y_{it-1},X_i) = 1 \\
&\phi_\theta^{(1,1)|(1,1)}((1,0),(1,1),Y_{it-1},X_i) = e^{-\left(\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)} \\
&\phi_\theta^{(1,1)|(1,1)}((0,1),(1,1),Y_{it-1},X_i) = e^{-\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1\right)} \\
&\phi_\theta^{(0,1)|(0,1)}((0,0),(1,1),Y_{it-1},X_i) = e^{-\left(\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)-\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1\right)} \\
&\qquad \phi_\theta^{(1,1)|(1,1)}(k,l,Y_{it-1},X_i) = 0, \quad \forall k \in \mathcal{Y}, \quad \forall l \in \mathcal{Y} \setminus \{(1,1)\}
\end{aligned}$$

This can be expressed compactly as:

$$\begin{aligned}
&\phi_\theta^{(1,1)|(1,1)}(Y_{it-1}^{t+1}, X_i) = Y_{1,it}Y_{2,it} \times \\
&e^{-(1-Y_{1,it+1})\left(\gamma_{11}(Y_{1,it-1}-1)+\gamma_{12}(Y_{2,it-1}-1)-\Delta X'_{1,it+1}\beta_1\right)-(1-Y_{2,it+1})\left(\gamma_{21}(Y_{1,it-1}-1)+\gamma_{22}(Y_{2,it-1}-1)-\Delta X'_{2,it+1}\beta_2\right)}
\end{aligned} \tag{21}$$

Given equations (18),(19),(20),(21), we can write succintly, $\forall k \in \mathcal{Y}$, when $M = 2$,

$$\begin{aligned}
\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = \mathbb{1}\{Y_{it} = k\} &e^{(Y_{1,it+1}-k_1)\left(\gamma_{11}(Y_{1,it-1}-k_1)+\gamma_{12}(Y_{2,it-1}-k_2)-\Delta X'_{1,it+1}\beta_1\right)} \times \\
&e^{(Y_{2,it+1}-k_1)\left(\gamma_{21}(Y_{1,it-1}-k_1)+\gamma_{22}(Y_{2,it-1}-k_2)-\Delta X_{2,it+1}\beta_2\right)}
\end{aligned}$$

In light of the expressions obtained for the bivariate VAR(1) case, we can conjecture that the transition function associated to $\pi_t^{k|k}(A_i, X_i), k \in \mathcal{Y}$ for the general M-variate case takes the form

$$\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i) = \mathbb{1}\{Y_{it} = k\} e^{\sum_{m=1}^M (Y_{m,it+1}-k_m)\left(\sum_{j=1}^M \gamma_{mj}(Y_{j,it-1}-k_j)-\Delta X'_{m,it+1}\beta_m\right)}$$

We verify that this is indeed correct by direct calculation.

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

Now, noting that

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

we finally get

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

which concludes the proof.

K Proof of Lemma 6

By definition, for $T \geq 3$, and for t, s such that $T-1 \geq t > s \geq 1$:

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

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 6. 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 10.

L Dynamic network formation with transitivity

Graham (2013) studies a variant of model (11) to describe network formation amongst groups of 3 individuals. This is a panel data setting where a large sample of many such groups and the evolution of their social ties are observed over $T = 3$ periods (4 counting the initial condition). Interactions are assumed undirected and modelled at the dyad level as:

$$
\begin{aligned}
D_{ijt} &= \mathbb{1}\left\{\gamma_0 D_{ijt-1} + \delta_0 R_{ijt-1} + A_{ij} - \epsilon_{ijt} \geq 0\right\} \quad t = 1, \ldots, T \\
R_{ijt-1} &= D_{ikt-1} D_{jkt-1}
\end{aligned}
\tag{22}
$$

where i, j, k denote the 3 different agents and $D_{ijt} \in \{0,1\}$ encodes the presence or absence of a link between agent i and agent j at time t. The network $D_0 \in \{0,1\}^3$ forms the initial condition. The parameter γ_0 captures state dependence while δ_0 captures transitivity in relationships, i.e the effect of sharing friends in common on the propensity to establish friendships. Finally, A_{ij} is an unrestricted dyad level fixed effect that could potentiall capture unobserved homophily and ϵ_{ijt} is a standard logistic shock, iid over time and individuals. While Graham (2013) establishes identification of (γ_0, δ_0) for $T = 3$ via a conditional likelihood approach in the spirit of Chamberlain (1985), one limitation of the model is the absence of other covariates, in particular time-specific effects. Controlling for such effects can be essential to adequately capture important variation in social dynamics: think about the persistent impact of Covid-19 on all types of social interactions.

A relevant extension is thus:

$$\begin{aligned} D_{ijt} &= \mathbb{1}\left\{\gamma_0 D_{ijt-1} + \delta_0 D_{ikt-1}D_{jkt-1} + X'_{ijt}\beta_0 + A_{ij} - \epsilon_{ijt} \geq 0\right\} \quad t = 1,\ldots,T \\ R_{ijt-1} &= D_{ikt-1}D_{jkt-1} \end{aligned} \tag{23}$$

Letting $\mathbb{D} = \{0,1\}^3$ denote the support of the network $D_t = (D_{ijt}, D_{ikt}, D_{jkt})$, it is straightforward to see that the results developed for the VAR(1) case can be repurposed to suit model (23) . For $T = 3$, an adaptation of Lemma 5 yields 8 possible transition functions given by:

$$\phi_\theta^{d|d}(D_3, D_2, D_1, X) = \mathbb{1}\{D_2 = d\} \exp\left(\sum_{i<j}(D_{ij3} - d_{ij2})[\gamma(D_{ij1} - d_{ij2}) - \Delta R_{ij1}\delta - \Delta X'_{ij2}\beta]\right), \quad d \in \mathbb{D}$$

An adaptation of Lemma 6 implies that we can construct another 8 transition functions given by

$$\zeta_\theta^{d|d}(D_3, D_2, D_1, D_0, X) = \mathbb{1}\{D_1 = d\} + \sum_{d' \in \mathbb{D}\setminus\{d\}} \omega_{2,1,d'}^{d|d}(\theta)\mathbb{1}\{D_1 = l\}\phi_\theta^{d|d}(D_3, D_2, D_2, X), \quad d \in \mathbb{D}$$

where

$$\begin{aligned} \mu_{ij,1}(\theta) &= \gamma D_{ij0} + \delta R_{ij0} + X'_{ij1}\beta \\ \kappa_{ij,2}^{d|d}(\theta) &= \gamma d_{ij} + \delta r_{ij} + X'_{ij3}\beta \\ \omega_{2,1,d'}^{d|d}(\theta) &= 1 - e^{\sum_{i<j}(d'_{ij} - d_{ij})\left[\kappa_{ij,2}^{d|d}(\theta) - \mu_{ij,1}(\theta)\right]} \end{aligned}$$

Therefore, for $T = 3$, 8 moment functions that all meaningfully depend on the model parameter are:

$$\psi_\theta^{d|d}(D_3, D_2, D_1, D_0, X) = \phi_\theta^{d|d}(D_3, D_2, D_1, X) - \zeta_\theta^{d|d}(D_3, D_2, D_1, D_0, X), \quad d \in \mathbb{D}$$

Their validity, in the sense of verifying equation (1), follows from the law of iterated expectations.

M Proof of Lemma 7

Let us first consider the case $C = 2$. The proposed functional form for the transition function $\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)$ implies that it is null when $Y_{it} \neq 0$. Hence

$$E\left[\phi_\theta^{0|0}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i\right] = \frac{e^{\sum_{l=0}^{2}\gamma_{0l}\mathbb{1}(Y_{it-1}=l)+X'_{i0t}\beta_0+A_{i0}}}{\sum_{j=0}^{2} e^{\sum_{l=0}^{2}\gamma_{jl}\mathbb{1}(Y_{it-1}=l)+X'_{ijt}\beta_j+A_{ij}}}\times$$

$$\left(\frac{e^{\gamma_{00}+X'_{i0t+1}\beta_0+A_{i0}}}{\sum_{j=0}^{2} e^{\gamma_{j0}+X'_{ijt+1}\beta_j+A_{ij}}}\phi_\theta^{0|0}(0,0,Y_{it-1},X_i)+\right.$$
$$\frac{e^{\gamma_{10}+X'_{i1t+1}\beta_1+A_{i1}}}{\sum_{j=0}^{2} e^{\gamma_{j0}+X'_{ijt+1}\beta_j+A_{ij}}}\phi_\theta^{0|0}(1,0,Y_{it-1},X_i)+$$
$$\left.\frac{e^{\gamma_{20}+X'_{i2t+1}\beta_2+A_{i2}}}{\sum_{j=0}^{2} e^{\gamma_{j0}+X'_{ijt+1}\beta_j+A_{ij}}}\phi_\theta^{0|0}(2,0,Y_{it-1},X_i)\right)$$

Thus, to obtain the transition probability $\pi_t^{0|0}(A_i, X_i) = \frac{e^{\gamma_{00}+X'_{i0t+1}\beta_j+A_{i0}}}{\sum_{j=0}^{2} e^{\gamma_{j0}+X'_{ijt+1}\beta_j+A_{ij}}}$, we must set

$$\phi_\theta^{0|0}(0,0,Y_{it-1},X_i) = 1$$
$$\phi_\theta^{0|0}(1,0,Y_{it-1},X_i) = \exp\left(\left(\sum_{j=0}^{2}\gamma_{1j}\mathbb{1}(Y_{it-1}=j) - \gamma_{10} - \Delta X'_{i1t+1}\beta_1\right) - \left(\sum_{j=0}^{2}\gamma_{0j}\mathbb{1}(Y_{it-1}=j) - \gamma_{00} - \Delta X'_{i0t+1}\beta_0\right)\right)$$
$$\phi_\theta^{0|0}(2,0,Y_{it-1},X_i) = \exp\left(\left(\sum_{j=0}^{2}\gamma_{2j}\mathbb{1}(Y_{it-1}=j) - \gamma_{20} - \Delta X'_{i2t+1}\beta_2\right) - \left(\sum_{j=0}^{2}\gamma_{0j}\mathbb{1}(Y_{it-1}=j) - \gamma_{00} - \Delta X'_{i0t+1}\beta_0\right)\right)$$

or more succinctly

$$\begin{aligned}\phi_\theta^{0|0}(Y_{it-1}^{t+1}, X_i) &= \mathbb{1}\{Y_{it}=0\}e^{\sum_{c=1}^{2}\mathbb{1}\{Y_{it+1}=c\}\left[\left(\sum_{j=0}^{2}\gamma_{cj}\mathbb{1}(Y_{it-1}=j)-\gamma_{c0}-\Delta X'_{ict+1}\beta_c\right)-\left(\sum_{j=0}^{2}\gamma_{0j}\mathbb{1}(Y_{it-1}=j)-\gamma_{00}-\Delta X'_{i0t+1}\beta_0\right)\right]}\\ &= \mathbb{1}\{Y_{it}=0\}e^{\sum_{c=1}^{2}\mathbb{1}\{Y_{it+1}=c\}\left[\left(\sum_{j=0}^{2}(\gamma_{cj}-\gamma_{0j})\mathbb{1}(Y_{it-1}=j)+\gamma_{00}-\gamma_{c0}+\Delta X'_{i0t+1}\beta_0-\Delta X'_{ict+1}\beta_c\right)\right]}\end{aligned}$$

By symmetry, it is clear that we also have:

$$\phi_\theta^{1|1}(Y_{it-1}^{t+1}, X_i) = \mathbb{1}\{Y_{it}=1\}e^{\sum_{c=0, c\neq 1}^{2}\mathbb{1}\{Y_{it+1}=c\}\left[\left(\sum_{j=0}^{2}(\gamma_{cj}-\gamma_{1j})\mathbb{1}(Y_{it-1}=j)+\gamma_{11}-\gamma_{c1}+\Delta X'_{i1t+1}\beta_1-\Delta X'_{ict+1}\beta_c\right)\right]}$$
$$\phi_\theta^{2|2}(Y_{it-1}^{t+1}, X_i) = \mathbb{1}\{Y_{it}=2\}e^{\sum_{c=0, c\neq 2}^{2}\mathbb{1}\{Y_{it+1}=c\}\left[\left(\sum_{j=0}^{2}(\gamma_{cj}-\gamma_{2j})\mathbb{1}(Y_{it-1}=j)+\gamma_{22}-\gamma_{c2}+\Delta X'_{i2t+1}\beta_2-\Delta X'_{ict+1}\beta_c\right)\right]}$$

In light of these expressions, one would naturally conjecture that for arbitrary $C \geq 2$ the expressions become for all $k \in \mathcal{Y}$:

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

We proceed to verify that this is correct by direct computation. We have:

$$E\left[\phi_\theta^{k|k}(Y_{it+1}, Y_{it}, Y_{it-1}, X_i)|Y_{i0}, Y_{i1}^{t-1}, X_i\right] = P(Y_{it} = k|Y_i^0, Y_{i1}^{t-1}, X_i, A_i)\times$$

$$\sum_{l\in\mathcal{Y}} P(Y_{it+1} = l|Y_i^0, Y_{i1}^{t-1}, Y_{it} = k, X_i, A_i)\phi_\theta^{k|k}(l, k, Y_{it-1}, X_i)$$

$$= \frac{e^{\sum_{c=0}^C \gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum_{j=0}^C e^{\sum_{c=0}^C \gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}}\times$$

$$\sum_{l\in\mathcal{Y}} \frac{e^{\gamma_{lk}+X'_{ilt+1}\beta_l+A_{il}}}{\sum_{j=0}^C e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}\phi_\theta^{k|k}(l, k, Y_{it-1}, X_i)$$

$$= \frac{e^{\sum_{c=0}^C \gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum_{j=0}^C e^{\sum_{c=0}^C \gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}}\times$$

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

$$= \frac{e^{\sum_{c=0}^C \gamma_{kc}\mathbb{1}(Y_{it-1}=c)+X'_{ikt}\beta_k+A_{ik}}}{\sum_{j=0}^C e^{\sum_{c=0}^C \gamma_{jc}\mathbb{1}(Y_{it-1}=c)+X'_{ijt}\beta_j+A_{ij}}} \times \frac{e^{\gamma_{kk}+X'_{ikt+1}\beta_k+A_{ik}}}{\sum_{j=0}^C e^{\gamma_{jk}+X'_{ijt+1}\beta_j+A_{ij}}}$$

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

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

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

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

which concludes the proof.

N Proof of Lemma 8

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

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

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 7. 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 9. Likewise, for all $k \in \mathcal{Y} \setminus \{0\}$,

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

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 $\mu_{k,s}(\theta)$ and follows from the law of iterated expectations and Lemma 7. The third line makes use of the definition of $\kappa^{k|k}_{c,t}(\theta)$ and $\omega^{k|k}_{t,s,l}(\theta)$. The fourth line uses the fact that $\kappa^{k|k}_{0,t}(\theta) = \mu_{0,s}(\theta) = 0$ due to the normalization $\gamma_{c0} = \gamma_{0c} = 0, A_{0c} = 0$ for all $c \in \mathcal{Y}$. The penultime line uses Appendix Lemma 9.

O Additional details for the Monte Carlo simulations

O.1 Expressions of the moment functions for the AR(3) with T=5

Using Theorem 1, Lemma 4 and Proposition 3, careful derivations give the following:

$$\psi^{0|0,1,1}_{\theta}(Y^5_{i-1}, Y^1_{i-2}, X_i) = -e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}(1-Y_{i1})Y_{i2}(1-Y_{i3})$$
$$+\left((1-e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta})\left(1-e^{-\gamma_1+\gamma_2+\gamma_3(1-Y_{i0})+X'_{i53}\beta}\right)e^{\gamma_1-\gamma_3-\Delta X'_{i5}\beta}-1\right)(1-Y_{i1})Y_{i2}Y_{i3}(1-Y_{i4})Y_{i5}$$
$$+\left((1-e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta})\left(1-e^{-\gamma_1+\gamma_2+\gamma_3(1-Y_{i0})+X'_{i53}\beta}\right)-1\right)(1-Y_{i1})Y_{i2}Y_{i3}(1-Y_{i4})(1-Y_{i5})$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})+\gamma_3(1-Y_{i-2})+X'_{i51}\beta}Y_{i1}(1-Y_{i2})+$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})+\gamma_3(1-Y_{i-2})+X'_{i51}\beta}(1-e^{-\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta})Y_{i1}Y_{i2}(1-Y_{i3})$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})+\gamma_3(1-Y_{i-2})+X'_{i51}\beta}(1-e^{-\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta})\left(1-e^{-\gamma_1+\gamma_3(1-Y_{i0})+X'_{i53}\beta}\right)e^{\gamma_1-\Delta X'_{i5}\beta}\times$$
$$Y_{i1}Y_{i2}Y_{i3}(1-Y_{i4})Y_{i5}$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})+\gamma_3(1-Y_{i-2})+X'_{i51}\beta}(1-e^{-\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X'_{i52}\beta})\left(1-e^{-\gamma_1+\gamma_3(1-Y_{i0})+X'_{i53}\beta}\right)\times$$
$$Y_{i1}Y_{i2}Y_{i3}(1-Y_{i4})(1-Y_{i5})$$
$$-(1-Y_{i1})Y_{i2}Y_{i3}Y_{i4}$$

$$\psi^{0|0,1,0}_{\theta}(Y^5_{i-1}, Y^1_{i-2}, X_i) = -\left(1-e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X'_{i,25}\beta}\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}Y_{i4}$$
$$-\left(1-e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X'_{i,25}\beta}\right)\left(1-\left(1-e^{\gamma_2-\gamma_3 Y_{i0}+X'_{i53}\beta}\right)e^{\gamma_1-\gamma_2-\Delta X'_{i5}\beta}\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}(1-Y_{i4})Y_{i5}$$
$$-\left(1-e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X'_{i,25}\beta}\right)e^{\gamma_2-\gamma_3 Y_{i0}+X'_{i53}\beta}(1-Y_{i1})(1-Y_{i2})Y_{i3}(1-Y_{i4})(1-Y_{i5})$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})-\gamma_3 Y_{i-2}+X'_{i,51}\beta}Y_{i1}(1-Y_{i2})(1-Y_{i3})$$
$$+e^{\gamma_1(1-Y_{i0})+\gamma_2(Y_{i0}-Y_{i-1})+\gamma_3(Y_{i-1}-Y_{i-2})+X'_{i,21}\beta}Y_{i1}(1-Y_{i2})Y_{i3}Y_{i4}$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})-\gamma_3 Y_{i-2}+X'_{i,51}\beta}\left[1-\left(1-e^{\gamma_1+\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X'_{i,25}\beta}\right)\left(1-\left(1-e^{-\gamma_3 Y_{i0}+X'_{i53}\beta}\right)e^{\gamma_1-\gamma_2+\gamma_3-\Delta X'_{i5}\beta}\right)\right]\times$$
$$Y_{i1}(1-Y_{i2})Y_{i3}(1-Y_{i4})Y_{i5}$$
$$+e^{-\gamma_1 Y_{i0}+\gamma_2(1-Y_{i-1})-\gamma_3 Y_{i-2}+X'_{i,51}\beta}\left[1-\left(1-e^{\gamma_1+\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X'_{i,25}\beta}\right)e^{-\gamma_3 Y_{i0}+X'_{i53}\beta}\right]Y_{i1}(1-Y_{i2})Y_{i3}(1-Y_{i4})(1-Y_{i5})$$
$$-(1-Y_{i1})Y_{i2}$$

$$\psi_\theta^{0|0,0,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) =$$

$$- (1 - Y_{i1})Y_{i2}Y_{i3}$$

$$- e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}(1 - Y_{i1})Y_{i2}(1 - Y_{i3})(1 - Y_{i4})(1 - Y_{i5})$$

$$+ \left(\left(1 - e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right) e^{\gamma_1+\gamma_3(Y_{i0}-1)+X_{i35}'\beta} - 1\right)(1 - Y_{i1})Y_{i2}(1 - Y_{i3})Y_{i4}$$

$$+ \left(\left(1 - e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1 - \left(1 - e^{\gamma_1+\gamma_3(Y_{i0}-1)+X_{i35}'\beta}\right)\left(1 - e^{\gamma_2-\gamma_3-\Delta X_{i5}'\beta}\right)\right) - 1\right) \times$$

$$(1 - Y_{i1})Y_{i2}(1 - Y_{i3})(1 - Y_{i4})Y_{i5}$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}+\gamma_3(1-Y_{i-2})+X_{i51}'\beta}Y_{i1}(1 - Y_{i2})$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}+\gamma_3(1-Y_{i-2})+X_{i51}'\beta}\left(1 - e^{-\gamma_1-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)Y_{i1}Y_{i2}(1 - Y_{i3})(1 - Y_{i4})(1 - Y_{i5})$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}+\gamma_3(1-Y_{i-2})+X_{i51}'\beta}\left(1 - e^{-\gamma_1-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)e^{\gamma_1+\gamma_2+\gamma_3(Y_{i0}-1)+X_{i35}'\beta}Y_{i1}Y_{i2}(1 - Y_{i3})Y_{i4}$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}+\gamma_3(1-Y_{i-2})+X_{i51}'\beta} \times$$

$$\left(1 - e^{-\gamma_1-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1 - \left(1 - e^{\gamma_1+\gamma_2+\gamma_3(Y_{i0}-1)+X_{i35}'\beta}\right)\left(1 - e^{\gamma_2-\Delta X_{i5}'\beta}\right)\right)Y_{i1}Y_{i2}(1 - Y_{i3})(1 - Y_{i4})Y_{i5}$$

$$\psi_\theta^{0|0,0,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) =$$

$$+ \left(e^{\gamma_2 Y_{i,0}+\gamma_3 Y_{i-1}+X_{25}'\beta} - 1\right)(1 - Y_{i1})(1 - Y_{i2})Y_{i3}$$

$$- \left(1 - e^{\gamma_2 Y_{i,0}+\gamma_3 Y_{i-1}+X_{25}'\beta}\right)\left(1 - e^{\gamma_3 Y_{i0}+X_{i35}'\beta}\right)(1 - Y_{i1})(1 - Y_{i2})(1 - Y_{i3})Y_{i4}$$

$$- \left(1 - e^{\gamma_2 Y_{i,0}+\gamma_3 Y_{i-1}+X_{25}'\beta}\right)\left(1 - e^{\gamma_3 Y_{i0}+X_{i35}'\beta}\right)\left(1 - e^{-\Delta X_{i5}'\beta}\right)(1 - Y_{i1})(1 - Y_{i2})(1 - Y_{i3})(1 - Y_{i4})Y_{i5}$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}-\gamma_3 Y_{i-2}+X_{i51}'\beta}Y_{i1}(1 - Y_{i2})(1 - Y_{i3})(1 - Y_{i4})(1 - Y_{i5})$$

$$+ e^{\gamma_1(1-Y_{i0})+\gamma_2(Y_{i0}-Y_{i-1})+\gamma_3(Y_{i-1}-Y_{i-2})+X_{i21}'\beta}Y_{i1}(1 - Y_{i2})Y_{i3}$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}-\gamma_3 Y_{i-2}+X_{i51}'\beta}\left(1 - \left(1 - e^{\gamma_1+\gamma_2 Y_{i,0}+\gamma_3 Y_{i-1}+X_{25}'\beta}\right)\left(1 - e^{\gamma_2+\gamma_3 Y_{i0}+X_{i35}'\beta}\right)\right)Y_{i1}(1 - Y_{i2})(1 - Y_{i3})Y_{i4}$$

$$+ e^{-\gamma_1 Y_{i0}-\gamma_2 Y_{i-1}-\gamma_3 Y_{i-2}+X_{i51}'\beta)} \times$$

$$\left(1 - \left(1 - e^{\gamma_1+\gamma_2 Y_{i,0}+\gamma_3 Y_{i-1}+X_{25}'\beta}\right)\left(1 - e^{\gamma_2+\gamma_3 Y_{i0}+X_{i35}'\beta}\right)\left(1 - e^{\gamma_3-\Delta X_{i5}'\beta}\right)\right)Y_{i1}(1 - Y_{i2})(1 - Y_{i3})(1 - Y_{i4})Y_{i5}$$

$$- (1 - Y_{i1})Y_{i2}$$

$$\psi_\theta^{1|1,0,0}(Y_{i-1}^5, Y_{i-2}^1, X_i)$$

$$= -e^{\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta} Y_{i1}(1-Y_{i2})Y_{i3}$$

$$+ \left(\left(1-e^{\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta}\right)\left(1-e^{-\gamma_1+\gamma_2+\gamma_3 Y_{i0}+X'_{i,35}\beta}\right)-1\right) Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}Y_{i5}$$

$$+ \left(\left(1-e^{\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta}\right)\left(1-e^{-\gamma_1+\gamma_2+\gamma_3 Y_{i0}+X'_{i,35}\beta}\right)e^{\gamma_1-\gamma_3+\Delta X'_{i5}\beta}-1\right) Y_{i1}(1-Y_{i2})(1-Y_{i3})Y_{i4}(1-Y_{i5})$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3 Y_{i-2}+X'_{i15}\beta}(1-Y_{i1})Y_{i2}$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3 Y_{i-2}+X'_{i15}\beta}\left(1-e^{-\gamma_1+\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta}\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3 Y_{i-2}+X'_{i15}\beta}\left(1-e^{-\gamma_1+\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta}\right)\left(1-e^{-\gamma_1+\gamma_3 Y_{i0}+X'_{i,35}\beta}\right)(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}Y_{i5}$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3 Y_{i-2}+X'_{i15}\beta}\times$$

$$\left(1-e^{-\gamma_1+\gamma_2 Y_{i0}+\gamma_3 Y_{i-1}+X'_{i25}\beta}\right)\left(1-e^{-\gamma_1+\gamma_3 Y_{i0}+X'_{i,35}\beta}\right)e^{\gamma_1+\Delta X'_{i5}\beta}(1-Y_{i1})(1-Y_{i2})(1-Y_{i3})Y_{i4}(1-Y_{i5})$$

$$- Y_{i1}(1-Y_{i2})(1-Y_{i3})(1-Y_{i4})$$

$$\psi_\theta^{1|1,0,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) =$$

$$+ \left(e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}-1\right) Y_{i1}Y_{i2}(1-Y_{i3})(1-Y_{i,4})$$

$$- \left(1-e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}\right)e^{\gamma_2+\gamma_3(Y_{i0}-1)+X'_{i,35}\beta} Y_{i1}Y_{i2}(1-Y_{i3})Y_{i4}Y_{i5}$$

$$- \left(1-e^{-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}\right)\left(1-\left(1-e^{\gamma_2+\gamma_3(Y_{i0}-1)+X'_{i,35}\beta}\right)e^{\gamma_1-\gamma_2+\Delta X'_{i5}\beta}\right) Y_{i1}Y_{i2}(1-Y_{i3})Y_{i4}(1-Y_{i5})$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3(Y_{i-2}-1)+X'_{i15}\beta}(1-Y_{i1})Y_{i2}Y_{i3}$$

$$+ e^{\gamma_1 Y_{i0}+\gamma_2(Y_{i-1}-Y_{i0})+\gamma_3(Y_{i-2}-Y_{i-1})+X'_{i12}\beta}(1-Y_{i1})Y_{i2}(1-Y_{i3})(1-Y_{i,4})$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3(Y_{i-2}-1)+X'_{i15}\beta}\left(1-\left(1-e^{\gamma_1-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}\right)e^{\gamma_3(Y_{i0}-1)+X'_{i,35}\beta}\right)(1-Y_{i1})Y_{i2}(1-Y_{i3})Y_{i4}Y_{i5}$$

$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2 Y_{i-1}+\gamma_3(Y_{i-2}-1)+X'_{i15}\beta}\times$$

$$\left(1-\left(1-e^{\gamma_1-\gamma_2 Y_{i0}+\gamma_3(1-Y_{i-1})+X'_{i52}\beta}\right)\left(1-\left(1-e^{\gamma_3(Y_{i0}-1)+X'_{i,35}\beta}\right)e^{\gamma_1-\gamma_2+\gamma_3+\Delta X'_{i5}\beta}\right)\right)(1-Y_{i1})Y_{i2}(1-Y_{i3})Y_{i4}(1-Y_{i5})$$

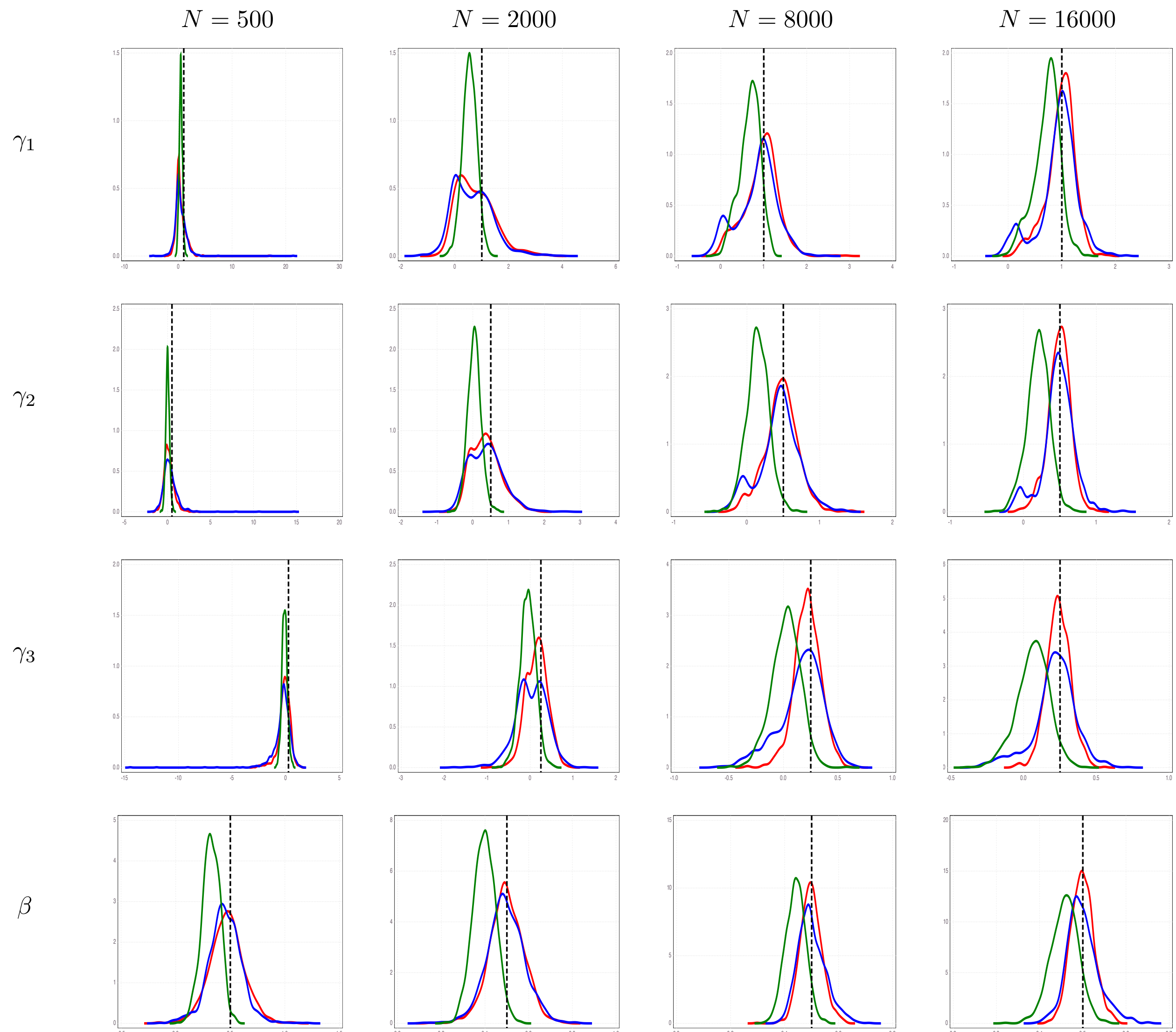
$$- Y_{i1}(1-Y_{i2})$$

$$\psi_\theta^{1|1,1,0}(Y_{i-1}^5, Y_{i-2}^1, X_i) =$$
$$- e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta} Y_{i1}(1-Y_{i2})Y_{i3}Y_{i4}Y_{i5}$$
$$+ \left(\left(1-e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta}\right) e^{\gamma_1-\gamma_3 Y_{i0}+X_{53}'\beta} - 1\right) Y_{i1}(1-Y_{i2})Y_{i3}(1-Y_{i4})$$
$$+ \left(\left(1-e^{\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta}\right)\left(1-\left(1-e^{\gamma_1-\gamma_3 Y_{i0}+X_{53}'\beta}\right)\left(1-e^{\gamma_2-\gamma_3+\Delta X_{i5}'\beta}\right)\right) - 1\right) Y_{i1}(1-Y_{i2})Y_{i3}Y_{i4}(1-Y_{i5})$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3 Y_{i-2}+X_{i15}'\beta}(1-Y_{i1})Y_{i2}$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3 Y_{i-2}+X_{i15}'\beta}\left(1-e^{-\gamma_1+\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta}\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}Y_{i4}Y_{i5}$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3 Y_{i-2}+X_{i15}'\beta}\left(1-e^{-\gamma_1+\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta}\right)e^{\gamma_1+\gamma_2-\gamma_3 Y_{i0}+X_{53}'\beta}(1-Y_{i1})(1-Y_{i2})Y_{i3}(1-Y_{i4})$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3 Y_{i-2}+X_{i15}'\beta}\times$$
$$\left(1-e^{-\gamma_1+\gamma_2(Y_{i0}-1)+\gamma_3 Y_{i-1}+X_{i25}'\beta}\right)\left(1-\left(1-e^{\gamma_1+\gamma_2-\gamma_3 Y_{i0}+X_{53}'\beta}\right)\left(1-e^{\gamma_2+\Delta X_{i5}'\beta}\right)\right)(1-Y_{i1})(1-Y_{i2})Y_{i3}Y_{i4}(1-Y_{i5})$$
$$- Y_{i1}(1-Y_{i2})(1-Y_{i3})$$

$$\psi_\theta^{1|1,1,1}(Y_{i-1}^5, Y_{i-2}^1, X_i) =$$
$$+ \left(e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta} - 1\right) Y_{i1}Y_{i2}(1-Y_{i3})$$
$$- \left(1-e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1-e^{\gamma_3(1-Y_{i0})+X_{53}'\beta}\right) Y_{i1}Y_{i2}Y_{i3}(1-Y_{i4})$$
$$- \left(1-e^{\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1-e^{\gamma_3(1-Y_{i0})+X_{53}'\beta}\right)\left(1-e^{\Delta X_{i5}'\beta}\right) Y_{i1}Y_{i2}Y_{i3}Y_{i4}(1-Y_{i5})$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3(Y_{i-2}-1)+X_{i15}'\beta}(1-Y_{i,1})Y_{i2}Y_{i3}Y_{i4}Y_{i5}$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3(Y_{i-2}-1)+X_{i15}'\beta}e^{\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}(1-Y_{i,1})Y_{i2}(1-Y_{i3})$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3(Y_{i-2}-1)+X_{i15}'\beta}\times$$
$$\left(1-\left(1-e^{\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1-e^{\gamma_2+\gamma_3(1-Y_{i0})+X_{53}'\beta}\right)\right)(1-Y_{i,1})Y_{i2}Y_{i3}(1-Y_{i4})$$
$$+ e^{\gamma_1(Y_{i0}-1)+\gamma_2(Y_{i-1}-1)+\gamma_3(Y_{i-2}-1)+X_{i15}'\beta}\times$$
$$\left(1-\left(1-e^{\gamma_1+\gamma_2(1-Y_{i0})+\gamma_3(1-Y_{i-1})+X_{i52}'\beta}\right)\left(1-e^{\gamma_2+\gamma_3(1-Y_{i0})+X_{53}'\beta}\right)\left(1-e^{\gamma_3+\Delta X_{i5}'\beta}\right)\right)(1-Y_{i,1})Y_{i2}Y_{i3}Y_{i4}(1-Y_{i5})$$
$$- Y_{i1}(1-Y_{i2})$$

O.2 Figures for the AR(3)

Figure 2: Densities of GMM estimators for the AR(3) with one regressor



Notes: The densities of estimates based on the first GMM estimator (i.e $\hat{\theta}^a$), the second GMM estimator (i.e $\hat{\theta}^b$) and the third GMM estimator (i.e $\hat{\theta}^c$) are indicated in green, blue and red respectively. Reported results are based on a 1000 replications of the DGP presented above with $\gamma_{01} = 1.0, \gamma_{02} = 0.5, \gamma_{03} = 0.25, \beta_0 = 0.5$. True parameter values are indicated with a vertical dashed line.

O.3 Expressions of the moment functions for the MAR(1) with $C = 3$ and $T = 3$

Here, we exceptionally use the notation $X_{ik,ts} = X_{ikt} - X_{iks}$ to keep manageable expressions. Appealing to Lemmas 7-8 and adapting Proposition 2 to the MAR(1) setting, careful calculations lead to:

$$
\begin{aligned}
\psi_\theta^{0|0}(Y_{i1}^3, Y_{i0}^1, X_i) = &\left(e^{X'_{i0;32}\beta_0 - X'_{i1;32}\beta_1} - 1\right)\mathbb{1}(Y_{i1} = 0)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 1\} \\
&+ \left(e^{X'_{i0;32}\beta_0 - X'_{i2;32}\beta_2} - 1\right)\mathbb{1}(Y_{i1} = 0)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 2\} \\
&+ e^{-\gamma_{11}\mathbb{1}\{Y_{i0}=1\} - \gamma_{12}\mathbb{1}\{Y_{i0}=2\} + X'_{i1;31}\beta_1 - X'_{i0;31}\beta_0}\mathbb{1}\{Y_{i1} = 1\}\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 0\} \\
&+ e^{\gamma_{11} - \gamma_{11}\mathbb{1}\{Y_{i0}=1\} - \gamma_{12}\mathbb{1}\{Y_{i0}=2\} + X'_{i1;21}\beta_1 - X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1} = 1)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 1\} \\
&+ e^{\gamma_{21} - \gamma_{11}\mathbb{1}\{Y_{i0}=1\} - \gamma_{12}\mathbb{1}\{Y_{i0}=2\} + X'_{i1;31}\beta_1 - X'_{i2;32}\beta_2 - X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1} = 1)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 2\} \\
&+ e^{-\gamma_{21}\mathbb{1}\{Y_{i0}=1\} - \gamma_{22}\mathbb{1}\{Y_{i0}=2\} + X'_{i2;31}\beta_2 - X'_{i0;31}\beta_0}\mathbb{1}\{Y_{i1} = 2\}\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 0\} \\
&+ e^{\gamma_{12} - \gamma_{21}\mathbb{1}\{Y_{i0}=1\} - \gamma_{22}\mathbb{1}\{Y_{i0}=2\} - X'_{i1;32}\beta_1 + X'_{i2;31}\beta_2 - X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1} = 2)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 1\} \\
&+ e^{-\gamma_{21}\mathbb{1}\{Y_{i0}=1\} + \gamma_{22} - \gamma_{22}\mathbb{1}\{Y_{i0}=2\} + X'_{i2;21}\beta_2 - X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1} = 2)\mathbb{1}\{Y_{i2} = 0\}\mathbb{1}\{Y_{i3} = 2\} \\
&- \mathbb{1}\{Y_{i1} = 0\}\mathbb{1}\{Y_{i2} = 1\} \\
&- \mathbb{1}\{Y_{i1} = 0\}\mathbb{1}\{Y_{i2} = 2\}
\end{aligned}
$$

$$
\begin{aligned}
\psi_\theta^{1|1}(Y_{i1}^3, Y_{i0}^1, X_i) = &+\left(e^{X'_{i1;32}\beta_1 - X'_{i0;32}\beta_0} - 1\right)\mathbb{1}(Y_{i1} = 1)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 0\} \\
&+ \left(e^{X'_{i1;32}\beta_1 - X'_{i2;32}\beta_2} - 1\right)\mathbb{1}(Y_{i1} = 1)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 2\} \\
&+ e^{-\gamma_{11} + \gamma_{11}\mathbb{1}(Y_{i0}=1) + \gamma_{12}\mathbb{1}(Y_{i0}=2) - X'_{i1;31}\beta_1 + X'_{i0;31}\beta_0}\mathbb{1}\{Y_{i1} = 0\}\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 1\} \\
&+ e^{\gamma_{11}\mathbb{1}(Y_{i0}=1) + \gamma_{12}\mathbb{1}(Y_{i0}=2) - X'_{i1;21}\beta_1 + X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1} = 0)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 0\} \\
&+ e^{-\gamma_{21} + \gamma_{11}\mathbb{1}(Y_{i0}=1) + \gamma_{12}\mathbb{1}(Y_{i0}=2) - X'_{i1;21}\beta_1 - X'_{i2;32}\beta_2 + X'_{i0;31}\beta_0}\mathbb{1}(Y_{i1} = 0)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 2\} \\
&+ e^{\gamma_{21} - \gamma_{11} + (\gamma_{11} - \gamma_{21})\mathbb{1}(Y_{i0}=1) + (\gamma_{12} - \gamma_{22})\mathbb{1}(Y_{i0}=2) - X'_{i1;31}\beta_1 + X'_{i2;31}\beta_2}\mathbb{1}(Y_{i1} = 2)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 1\} \\
&+ e^{\gamma_{21} - \gamma_{12} + (\gamma_{11} - \gamma_{21})\mathbb{1}(Y_{i0}=1) + (\gamma_{12} - \gamma_{22})\mathbb{1}(Y_{i0}=2) - X'_{i1;21}\beta_1 + X'_{i2;31}\beta_2 - X'_{i0;32}\beta_0}\mathbb{1}(Y_{i1} = 2)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 0\} \\
&+ e^{\gamma_{22} - \gamma_{12} + (\gamma_{11} - \gamma_{21})\mathbb{1}(Y_{i0}=1) + (\gamma_{12} - \gamma_{22})\mathbb{1}(Y_{i0}=2) - X'_{i1;21}\beta_1 + X'_{i2;21}\beta_2}\mathbb{1}(Y_{i1} = 2)\mathbb{1}\{Y_{i2} = 1\}\mathbb{1}\{Y_{i3} = 2\} \\
&- \mathbb{1}\{Y_{i1} = 1\}\mathbb{1}\{Y_{i2} = 0\} \\
&- \mathbb{1}\{Y_{i1} = 1\}\mathbb{1}\{Y_{i2} = 2\}
\end{aligned}
$$

$$
\begin{aligned}
\psi_\theta^{2|2}(Y_{i1}^3, Y_{i0}^1, X_i) &= \left(e^{X'_{i2;32}\beta_2 - X'_{i0;32}\beta_0} - 1\right)\mathbb{1}(Y_{i1}=2)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=0\} \\
&+ \left(e^{X'_{i2;32}\beta_2 - X'_{i1;32}\beta_1} - 1\right)\mathbb{1}(Y_{i1}=2)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=1\} \\
&+ e^{-\gamma_{22}+\gamma_{21}\mathbb{1}(Y_{i0}=1)+\gamma_{22}\mathbb{1}(Y_{i0}=2)-X'_{i2;31}\beta_2+X'_{i0;31}\beta_0}\mathbb{1}\{Y_{i1}=0\}\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=2\} \\
&+ e^{\gamma_{21}\mathbb{1}(Y_{i0}=1)+\gamma_{22}\mathbb{1}(Y_{i0}=2)-X'_{i2;21}\beta_2+X'_{i0;21}\beta_0}\mathbb{1}(Y_{i1}=0)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=0\} \\
&+ e^{-\gamma_{12}+\gamma_{21}\mathbb{1}(Y_{i0}=1)+\gamma_{22}\mathbb{1}(Y_{i0}=2)-X'_{i2;21}\beta_2-X'_{i1;32}\beta_1+X'_{i0;31}\beta_0}\mathbb{1}(Y_{i1}=0)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=1\} \\
&+ e^{\gamma_{12}-\gamma_{22}+(\gamma_{21}-\gamma_{11})\mathbb{1}(Y_{i0}=1)+(\gamma_{22}-\gamma_{12})\mathbb{1}(Y_{i0}=2)+X'_{i1;31}\beta_1-X'_{i2;31}\beta_2}\mathbb{1}\{Y_{i1}=1\}\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=2\} \\
&+ e^{\gamma_{12}-\gamma_{21}+(\gamma_{21}-\gamma_{11})\mathbb{1}(Y_{i0}=1)+(\gamma_{22}-\gamma_{12})\mathbb{1}(Y_{i0}=2)+X'_{i1;31}\beta_1-X'_{i2;21}\beta_2-X'_{i0;32}\beta_0}\mathbb{1}(Y_{i1}=1)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=0\} \\
&+ e^{\gamma_{11}-\gamma_{21}+(\gamma_{21}-\gamma_{11})\mathbb{1}(Y_{i0}=1)+(\gamma_{22}-\gamma_{12})\mathbb{1}(Y_{i0}=2)+X'_{i1;21}\beta_1-X'_{i2;21}\beta_2}\mathbb{1}(Y_{i1}=1)\mathbb{1}\{Y_{i2}=2\}\mathbb{1}\{Y_{i3}=1\} \\
&- \mathbb{1}\{Y_{i1}=2\}\mathbb{1}\{Y_{i2}=0\} \\
&- \mathbb{1}\{Y_{i1}=2\}\mathbb{1}\{Y_{i2}=1\}
\end{aligned}
$$