

Linear estimations of dynamic fixed effects logit models only with time effects*

Yoshitsugu Kitazawa[†]
Kyushu Sangyo University

March 24, 2026

Abstract

This paper proposes linear estimation methods for dynamic fixed effects logit models only with time effects (i.e., those only with time dummies and only with time trends). The linear estimators point-identify transformations of parameters of interest for the models if five or more time periods are provided and then point-identify the parameters of interest. What it boils down to is that root- N consistent estimations are attainable for these models. Monte Carlo results corroborate this conclusion.

Keywords: dynamic panel logit models, fixed effects, time dummies, time trends, point-identification, root- N consistent estimators, Monte Carlo experiments

JEL Codes: C23, C25, C26

1 INTRODUCTION

Honoré and Tamer (2006) suggest that neither the dynamic fixed effects binary choice model only with time trends nor that only with time dummies is point-identified after meticulously inspecting their calculation results from a linear programming method using probit models. However, instead, as shown from now, the logit models included in them are point-identified if five consecutive time periods are provided. Disappointingly, they overlook this significant discovery without showing the calculation results for the case of five time periods and without showing the results from using logit specifications.

*Discussion Paper Series, Faculty of Economics, Kyushu Sangyo University Discussion Paper, March 2026, No. 87. Corrections and Revisions:

http://www.ip.kyusan-u.ac.jp/J/kitazawa/ERRATA/errata_linear.html

[†]Correspondence: Yoshitsugu Kitazawa, Faculty of Economics, Kyushu Sangyo University, Matsukadai 2-3-1, Higashi-ku, Fukuoka, Japan. E-mail: kitazawa@ip.kyusan-u.ac.jp

The matching approach by Honoré and Kyriazidou (2000) is not essential for achieving point-identification for dynamic fixed effects logit models. To show this, we propose (globally) root- N consistent estimators for both models only with time dummies and only with time trends after proving point-identification of these models.

For dynamic fixed effects logit models with strictly exogenous continuous explanatory variables and/or time effects, Kitazawa (2022) is the first to obtain the valid moment conditions, leveraging the properties of hyperbolic functions.¹ After the transformations of these moment conditions give birth to linear asymptotically normal and root- N consistent estimators for the transformations of parameters of interest in the models only with time effects, asymptotically normal and root- N consistent estimators are obtained for the parameters of interest without nonlinear optimization.

In retrospect, no estimators are heretofore proposed for point-identifying the parameters of interest in the dynamic fixed effects logit models without imposing some additional restrictions, except in the simple models²: Honoré and Kyriazidou (2000) propose a kernel-using estimator requiring that first-differences of the explanatory variables at a specific time period are continuously distributed with support near zero, which excludes the time effects and explanatory variables such as age and spouse income based on the age-based remuneration system, etc.³ Honoré and Weidner (2025) show that the usage of the moment conditions mentioned in previous paragraph point-identifies parameters of interest for the model with strictly exogenous explanatory variables if a special dataset is used. Although Dobronyi et al. (2021) provide specific numerical examples of point-identification for the models with time effects, where they use the moment inequality conditions they propose in addition to the moment (equality) conditions in previous paragraph, it is unknown whether the models are generically point-identified or not.⁴

Although it remains an open question whether the models with strictly exogenous continuous explanatory variables are point-identified or not, this paper radically sorts out the controversy as to whether the models only with time effects are point-identified or not.

The rest of the paper is as follows: section 2 constructs systems of simultaneous equations consisting of the transformed moment conditions to obtain solutions of the trans-

¹Honoré and Weidner (2025), Dobronyi et al. (2021), and Dano (2023) also construct the equivalent moment conditions in different ways: the functional differencing approach proposed by Bonhomme (2012), the formulation of the full likelihood, and the method based on transition probabilities, respectively.

²Chamberlain (1985) shows that the model with neither explanatory variables nor time effects is point-identified. Further, Honoré and Kyriazidou (2000) show that the model only with binary explanatory variables is point-identified as well. They propose root- N consistent estimators for the respective models.

³These variables could be used in studies on the determinants of female labor force participation.

⁴The 2025 version of Dobronyi et al. (2021), where T.M. Russell joins as one of the coauthors, definitely states in footnote 34 that it is not known whether the time trend model is generically point-identified or not when four time periods (incorporating the initial observations) are provided. No statements are found for the cases of five or more time periods. To the best of author's knowledge, no literatures and documents are available concerning proofs of point-identification of the time trend and time dummies models, although the analogy stated in Remark 8 in the supplement of the 2025 version of Dano (2023) claims that the AR(2) (autoregressive of order 2) time trend model will be point-identified when six time periods are provided: we can fetch the supplement from Dano's website.

formed parameters uniquely. Then, values of the original parameters are solved from their solutions uniquely. Section 3 presents Monte Carlo illustrations giving evidence for asymptotic normality and root- N consistency of the estimators proposed in section 2. Finally, section 4 clears up the misunderstanding surmised from the results obtained by Honoré and Tamer (2006) and concludes the paper.

2 LINEAR MOMENT CONDITIONS

Throughout the paper, the subscripts i ($i = 1, \dots, N$) and t ($t = 1, \dots, T$) denote the individual and time period, respectively. It is assumed that the number of individuals $N \rightarrow \infty$, while the number of time periods T is fixed. We use any five consecutive time periods within T periods to estimate the models.

2.1 Model with time dummies

The model only with time dummies (TD_t for $2 \leq t \leq T$) that gives birth to the binary dependent variable y_{it} is as follows:

$$y_{it} = p(\eta_i, y_{i,t-1}, TD_t) + v_{it}, \quad \text{for } 2 \leq t \leq T,$$

$$\text{with } E[v_{it} \mid \eta_i, y_{i1}, v_i^{t-1}] = 0,$$

where the logit probability $p(\eta_i, y_{i,t-1}, TD_t) = \exp(\eta_i + \gamma y_{i,t-1} + TD_t) / (1 + \exp(\eta_i + \gamma y_{i,t-1} + TD_t))$ with η_i and γ being the fixed effect and state dependence parameter respectively, with which $y_{it} = 1$. v_{it} is the disturbance, y_{i1} is the initial value of the binary dependent variable, and $v_i^{t-1} = (v_{i1}, \dots, v_{i,t-1})$ with v_{i1} being empty. We postulate that in this model, the original parameters (i.e., parameters of interest to be estimated) are γ , ΔTD_t , and ΔTD_{t+1} for time t , where $\Delta TD_t = TD_t - TD_{t-1}$.

Kitazawa (2022) proposes two types of moment conditions for the above model. We have those based on the g -form and h -form respectively as follows:

$$E[\hbar U_{it}^- \mid \eta_i, y_{i1}, v_i^{t-2}] = 0, \quad \text{for } 3 \leq t \leq T - 1,$$

$$\text{with } \hbar U_{it}^- = U_{it}^- - y_{i,t-1}$$

$$+ ((\Psi_t - \Phi_t)y_{i,t-2} + \Phi_t)$$

$$\times ((U_{it}^- - y_{i,t-1}) - 2U_{it}^-(1 - y_{i,t-1})),$$

$$U_{it}^- = y_{it} + (1 - y_{it})y_{i,t+1} - (1 - y_{it})y_{i,t+1}\phi_{t+1}^{-1}$$

$$- \delta y_{i,t-1}(1 - y_{it})y_{i,t+1}\phi_{t+1}^{-1},$$

and

$$E[\hbar \Upsilon_{it}^- \mid \eta_i, y_{i1}, v_i^{t-2}] = 0, \quad \text{for } 3 \leq t \leq T - 1,$$

$$\text{with } \hbar \Upsilon_{it}^- = \Upsilon_{it}^- - y_{i,t-1}$$

$$+ ((\Psi_t^* - \Phi_t^*)(1 - y_{i,t-2}) + \Phi_t^*)$$

$$\begin{aligned} & \times ((\Upsilon_{it}^- - y_{i,t-1}) - 2(\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1}), \\ \Upsilon_{it}^- &= y_{it}y_{i,t+1} + y_{it}(1 - y_{i,t+1})\phi_{t+1} \\ & + \delta(1 - y_{i,t-1})y_{it}(1 - y_{i,t+1})\phi_{t+1}, \end{aligned}$$

where putting $\delta = \exp(\gamma) - 1$, $\phi_t = \exp(\Delta T D_t)$, and $\phi_t^{-1} = 1/\phi_t$,

$$\begin{aligned} \Psi_t &= \frac{\phi_t\phi_{t+1} - (\delta + 1)}{\phi_t\phi_{t+1} + (\delta + 1)}; \quad \Phi_t = \frac{\phi_t\phi_{t+1} - 1}{\phi_t\phi_{t+1} + 1}; \\ \Psi_t^* &= \frac{\phi_t^{-1}\phi_{t+1}^{-1} - (\delta + 1)}{\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1)}; \quad \Phi_t^* = \frac{\phi_t^{-1}\phi_{t+1}^{-1} - 1}{\phi_t^{-1}\phi_{t+1}^{-1} + 1}. \end{aligned}$$

By putting

$$\begin{aligned} \Theta_{it}^{(1)} &= (1 - y_{i,t-1})(y_{it} + (1 - y_{it})y_{i,t+1}), \\ \Theta_{it}^{(2)} &= -(1 - y_{i,t-1})(1 - y_{it})y_{i,t+1}, \\ \Theta_{it}^{(3)} &= y_{i,t-1}(y_{it} + (1 - y_{it})y_{i,t+1} - y_{i,t-1}), \\ \Theta_{it}^{(4)} &= -y_{i,t-1}(1 - y_{it})y_{i,t+1}, \\ \Xi_{it}^{(1)} &= y_{i,t-1}(y_{it}y_{i,t+1} - y_{i,t-1}), \\ \Xi_{it}^{(2)} &= y_{i,t-1}y_{it}(1 - y_{i,t+1}), \\ \Xi_{it}^{(3)} &= (1 - y_{i,t-1})y_{it}y_{i,t+1}, \\ \Xi_{it}^{(4)} &= (1 - y_{i,t-1})y_{it}(1 - y_{i,t+1}), \end{aligned}$$

we write

$$\begin{aligned} \iota_{it} &= (1/2)(\phi_t\phi_{t+1} + 1)(1 - y_{i,t-2})\hbar U_{it}^- \\ &= (1 - y_{i,t-2})(\Theta_{it}^{(1)} + \alpha_b^{(A)}\Theta_{it}^{(2)} + \alpha_c^{(A)}\Theta_{it}^{(3)} + \alpha_d^{(A)}\Theta_{it}^{(4)}), \\ \iota_{it}^* &= (1/2)(\phi_t\phi_{t+1} + (\delta + 1))(1 + \delta)^{-1}y_{i,t-2}\hbar U_{it}^- \\ &= y_{i,t-2}(\Theta_{it}^{(1)} + \alpha_b^{(A)}\Theta_{it}^{(2)} + \alpha_g^{(A)}\Theta_{it}^{(3)} + \alpha_a^{(A)}\Theta_{it}^{(4)}), \\ \kappa_{it} &= (1/2)(\phi_t^{-1}\phi_{t+1}^{-1} + 1)\phi_t(1 + \delta)^{-1}y_{i,t-2}\hbar \Upsilon_{it}^- \\ &= y_{i,t-2}(\alpha_e^{(A)}\Xi_{it}^{(1)} + \alpha_g^{(A)}\Xi_{it}^{(2)} + \alpha_f^{(A)}\Xi_{it}^{(3)} + \Xi_{it}^{(4)}), \\ \kappa_{it}^* &= (1/2)(\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))\phi_t(1 + \delta)^{-1}(1 - y_{i,t-2})\hbar \Upsilon_{it}^- \\ &= (1 - y_{i,t-2})(\alpha_a^{(A)}\Xi_{it}^{(1)} + \alpha_c^{(A)}\Xi_{it}^{(2)} + \alpha_f^{(A)}\Xi_{it}^{(3)} + \Xi_{it}^{(4)}), \end{aligned}$$

where the transformed parameters $\alpha_a^{(A)} = \phi_t$, $\alpha_b^{(A)} = \phi_{t+1}^{-1}$, $\alpha_c^{(A)} = \phi_t\phi_{t+1}$, $\alpha_d^{(A)} = \phi_t(1 + \delta)$, $\alpha_e^{(A)} = \phi_t(1 + \delta)^{-1}$, $\alpha_f^{(A)} = \phi_{t+1}^{-1}(1 + \delta)^{-1}$, and $\alpha_g^{(A)} = \phi_t\phi_{t+1}(1 + \delta)^{-1}$. Therefore, for time t , we have the following eight unconditional moment conditions linear in the seven parameters above:

$$\mathbf{E}[(1 \ y_{i,t-3})' \otimes (\iota_{it} \ \iota_{it}^* \ \kappa_{it} \ \kappa_{it}^*)'] = \mathbf{0}, \quad (1)$$

where $\mathbf{0}$ is a zero vector. The derivation of ι_{it} , ι_{it}^* , κ_{it} , and κ_{it}^* is provided in appendix A. A direct computation with Maxima (2025) suggests that the moment functions composing (1) are linearly independent for $(y_{i,t-3}, \dots, y_{i,t+1}) \in \{0, 1\}^5$ and almost all values of the parameters γ , ΔTD_t , and ΔTD_{t+1} .⁵

We can remove any one moment condition from the moment conditions (1) to produce an unique just-identified moment estimator for the set of the seven parameters $\alpha_a^{(A)}$, $\alpha_b^{(A)}$, $\alpha_c^{(A)}$, $\alpha_d^{(A)}$, $\alpha_e^{(A)}$, $\alpha_f^{(A)}$, and $\alpha_g^{(A)}$ as long as the rank condition is satisfied in constructing the estimator; hereafter we call it the *A* estimator (see details in appendices B and C). Then, we can obtain some unique estimators (without numerical optimization) for the set of the original parameters γ , ΔTD_t , and ΔTD_{t+1} , utilizing some relationships among the seven parameters (see details in appendix D). It should be noted that all these estimators are asymptotically normal and root- N consistent.

Of even more greater interest is the fact that using the moment conditions (1) without $E[(1 \ y_{i,t-3})' \ \kappa_{it}] = 0$ and (1) without $E[(1 \ y_{i,t-3})' \ \iota_{it}] = 0$ can give birth to unique just-identified moment estimators for the sets of the following six parameters $\alpha_a^{(A)}$, $\alpha_b^{(A)}$, $\alpha_c^{(A)}$, $\alpha_d^{(A)}$, $\alpha_f^{(A)}$, and $\alpha_g^{(A)}$ and the following six parameters $\alpha_a^{(A)}$, $\alpha_b^{(A)}$, $\alpha_c^{(A)}$, $\alpha_e^{(A)}$, $\alpha_f^{(A)}$, and $\alpha_g^{(A)}$, respectively; these estimators also count among the *A* estimators (see details in appendices B and C). Using these estimators, we can obtain asymptotically normal and root- N consistent estimators for γ , ΔTD_t , and ΔTD_{t+1} , as well (see details in appendix D).

Alternatively, using

$$\begin{aligned}\lambda_{it} &= (1/2)(\phi_t \phi_{t+1} + 1)\phi_t^{-1}(1 + \delta)^{-1}(1 - y_{i,t-2})\hbar U_{it}^- \\ &= (1 - y_{i,t-2})(\alpha_e^{(B)} \Theta_{it}^{(1)} + \alpha_g^{(B)} \Theta_{it}^{(2)} + \alpha_f^{(B)} \Theta_{it}^{(3)} + \Theta_{it}^{(4)}), \\ \lambda_{it}^* &= (1/2)(\phi_t \phi_{t+1} + (\delta + 1))\phi_t^{-1}(1 + \delta)^{-1}y_{i,t-2}\hbar U_{it}^- \\ &= y_{i,t-2}(\alpha_a^{(B)} \Theta_{it}^{(1)} + \alpha_c^{(B)} \Theta_{it}^{(2)} + \alpha_f^{(B)} \Theta_{it}^{(3)} + \Theta_{it}^{(4)}), \\ \mu_{it} &= (1/2)(\phi_t^{-1}\phi_{t+1}^{-1} + 1)y_{i,t-2}\hbar \Upsilon_{it}^- \\ &= y_{i,t-2}(\Xi_{it}^{(1)} + \alpha_b^{(B)} \Xi_{it}^{(2)} + \alpha_c^{(B)} \Xi_{it}^{(3)} + \alpha_d^{(B)} \Xi_{it}^{(4)}), \\ \mu_{it}^* &= (1/2)(\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))(1 + \delta)^{-1}(1 - y_{i,t-2})\hbar \Upsilon_{it}^- \\ &= (1 - y_{i,t-2})(\Xi_{it}^{(1)} + \alpha_b^{(B)} \Xi_{it}^{(2)} + \alpha_g^{(B)} \Xi_{it}^{(3)} + \alpha_a^{(B)} \Xi_{it}^{(4)}),\end{aligned}$$

where the transformed parameters $\alpha_a^{(B)} = \phi_t^{-1}$, $\alpha_b^{(B)} = \phi_{t+1}$, $\alpha_c^{(B)} = \phi_t^{-1}\phi_{t+1}^{-1}$, $\alpha_d^{(B)} = \phi_t^{-1}(1 + \delta)$, $\alpha_e^{(B)} = \phi_t^{-1}(1 + \delta)^{-1}$, $\alpha_f^{(B)} = \phi_{t+1}(1 + \delta)^{-1}$, and $\alpha_g^{(B)} = \phi_t^{-1}\phi_{t+1}^{-1}(1 + \delta)^{-1}$, we have the following eight unconditional moment conditions linear in the seven parameters above:

$$E[(1 \ y_{i,t-3})' \otimes (\lambda_{it} \ \lambda_{it}^* \ \mu_{it} \ \mu_{it}^*)'] = \mathbf{0}. \quad (2)$$

⁵The discussion on the linear dependence of the moment functions is also conducted in Honoré and Weidner (2025), Krueger (2020), Dobronyi et al. (2021), and Dano (2023), including strictly exogenous continuous explanatory variables and longer lagged dependent variables.

The derivation of λ_{it} , λ_{it}^* , μ_{it} , and μ_{it}^* is provided in appendix A. Given that these moment conditions (2) are a variant of (1), we suggest that the moment functions composing (2) are linearly independent as well as those composing (1). A calculation with Maxima (2025) also suggests this.

For these moment conditions (2), we can obtain some asymptotically normal and root- N consistent estimators for the the original parameters γ , ΔTD_t , and ΔTD_{t+1} in the same manner as for the moment conditions (1), where three types of the B estimators are defined: one using the moment conditions (2) with any one moment condition removed for the set of the seven parameters $\alpha_a^{(B)}$, $\alpha_b^{(B)}$, $\alpha_c^{(B)}$, $\alpha_d^{(B)}$, $\alpha_e^{(B)}$, $\alpha_f^{(B)}$, and $\alpha_g^{(B)}$, one using (2) without $E[(1 - y_{i,t-3})' \lambda_{it}] = 0$ for the set of the six parameters $\alpha_a^{(B)}$, $\alpha_b^{(B)}$, $\alpha_c^{(B)}$, $\alpha_d^{(B)}$, $\alpha_f^{(B)}$, and $\alpha_g^{(B)}$, and one using (2) without $E[(1 - y_{i,t-3})' \mu_{it}] = 0$ for the set of the six parameters $\alpha_a^{(B)}$, $\alpha_b^{(B)}$, $\alpha_c^{(B)}$, $\alpha_e^{(B)}$, $\alpha_f^{(B)}$, and $\alpha_g^{(B)}$ (see details in appendices B, C, and D).

In addition, using both the moment conditions available in (1) or (2) dated $t - 1$ and some of the above estimators for the transformed parameters, we can point-identify the original parameter ΔTD_{t-1} and obtain asymptotically normal and root- N consistent two-step estimators for it (see details in appendix E).

If we estimate all the first-differences of time dummies available within T periods (i.e., $\Delta TD_3, \dots, \Delta TD_T$), we should use the aforementioned estimators for any sets of five consecutive time periods.

2.2 Model with time trends

The model only with time trends ($\varphi(t - \tau)$ for $2 \leq t \leq T$ where φ and τ are parameters) is as follows:

$$y_{it} = p(\eta_i, y_{i,t-1}, t) + v_{it}, \quad \text{for } 2 \leq t \leq T,$$

$$\text{with } E[v_{it} \mid \eta_i, y_{i1}, v_i^{t-1}] = 0,$$

where the logit probability $p(\eta_i, y_{i,t-1}, t) = \exp(\eta_i + \gamma y_{i,t-1} + \varphi(t - \tau)) / (1 + \exp(\eta_i + \gamma y_{i,t-1} + \varphi(t - \tau)))$ with which $y_{it} = 1$ and we inherit the notation from previous subsection other than the time trends. In this model, the original parameters are γ and φ .

With $\phi = \phi_t = \phi_{t+1} = \exp(\varphi)$ in $\hbar U_{it}^-$ and $\hbar \Upsilon_{it}^-$, we define

$$\begin{aligned} \varrho_{it} &= (1/2)(\phi^2 + 1)(1 - y_{i,t-2})\hbar U_{it}^- \\ &= (1 - y_{i,t-2})(\Theta_{it}^{(1)} + \alpha_b^{(C)} \Theta_{it}^{(2)} + \alpha_c^{(C)} \Theta_{it}^{(3)} + \alpha_e^{(C)} \Theta_{it}^{(4)}), \\ \varrho_{it}^* &= (1/2)(\phi^2 + (\delta + 1))\phi^{-1}(1 + \delta)^{-1}y_{i,t-2}\hbar U_{it}^- \\ &= y_{i,t-2}(\alpha_b^{(C)} \Theta_{it}^{(1)} + \alpha_d^{(C)} \Theta_{it}^{(2)} + \alpha_f^{(C)} \Theta_{it}^{(3)} + \Theta_{it}^{(4)}), \\ \varsigma_{it} &= (1/2)(\phi^{-2} + 1)y_{i,t-2}\hbar \Upsilon_{it}^- \\ &= y_{i,t-2}(\Xi_{it}^{(1)} + \alpha_a^{(C)} \Xi_{it}^{(2)} + \alpha_d^{(C)} \Xi_{it}^{(3)} + \alpha_g^{(C)} \Xi_{it}^{(4)}), \\ \varsigma_{it}^* &= (1/2)(\phi^{-2} + (\delta + 1))\phi(1 + \delta)^{-1}(1 - y_{i,t-2})\hbar \Upsilon_{it}^- \\ &= (1 - y_{i,t-2})(\alpha_a^{(C)} \Xi_{it}^{(1)} + \alpha_c^{(C)} \Xi_{it}^{(2)} + \alpha_h^{(C)} \Xi_{it}^{(3)} + \Xi_{it}^{(4)}), \end{aligned}$$

where the transformed parameters $\alpha_a^{(C)} = \phi$, $\alpha_b^{(C)} = \phi^{-1}$, $\alpha_c^{(C)} = \phi^2$, $\alpha_d^{(C)} = \phi^{-2}$, $\alpha_e^{(C)} = \phi(1 + \delta)$, $\alpha_f^{(C)} = \phi(1 + \delta)^{-1}$, $\alpha_g^{(C)} = \phi^{-1}(1 + \delta)$, and $\alpha_h^{(C)} = \phi^{-1}(1 + \delta)^{-1}$. Therefore, for time t , we have the following eight unconditional moment conditions linear in the eight parameters above:

$$\mathbb{E}[(\varrho_{it} \ \varrho_{it}^* \ \varsigma_{it} \ \varsigma_{it}^* \ \varrho_{i,t-1} \ \varrho_{i,t-1}^* \ \varsigma_{i,t-1} \ \varsigma_{i,t-1}^*)'] = \mathbf{0}. \quad (3)$$

The derivation of ϱ_{it} , ϱ_{it}^* , ς_{it} , and ς_{it}^* is provided in appendix A. A calculation similar to the case regarding (1) suggests the linear independence among the moment functions composing (3) for almost all values of the parameters γ and φ .

The just-identified estimator for the set of the eight parameters $\alpha_a^{(C)}$, $\alpha_b^{(C)}$, $\alpha_c^{(C)}$, $\alpha_d^{(C)}$, $\alpha_e^{(C)}$, $\alpha_f^{(C)}$, $\alpha_g^{(C)}$, and $\alpha_h^{(C)}$ is constructed using all the moment conditions (3) as long as the rank condition is satisfied, similar to those in previous subsection; hereafter we call it the C estimator (see details in appendix B). Then, we can obtain some unique estimators for the original parameters γ and φ (see details in appendix D). All these estimators are asymptotically normal and root- N consistent.

In addition, the A and B estimators proposed in previous subsection point-identify the original parameters γ and φ using some relationships among the transformed parameters, as well. Obviously, their estimators are asymptotically normal and root- N consistent.

3 MONTE CARLO

In this section, the finite sample behavior of estimators using the moment conditions proposed in previous section is investigated with Monte Carlo experiments. For the two models in previous section, the experiments are designed with the number of replications $R = 2500$, the number of generated time periods $T = 8$, and the cross-sectional sizes $N = (10 \text{ million}, 50 \text{ million}, 100 \text{ million})$. The first three cross-sections are discarded, while the last five ones are used for the estimations. The experiments are carried out with a program written in Julia developed by Bezanson et al. (2017).⁶

We present the tables displaying the Monte Carlo results, including mean (Monte Carlo mean), sd (Monte Carlo standard deviation), se (Monte Carlo mean of standard errors), bias (Monte Carlo bias), and rmse (Monte Carlo root mean squared error).

In each replication, estimates of the original parameters and those from using two-step estimations are calculated according to the illustrations in appendices D and E, respectively.

3.1 Model with time dummies

The DGP (data generating process) is as follows:

$$y_{it} = \begin{cases} 1 & \text{if } p(\eta_i, y_{i,t-1}, TD_t) > \zeta_{it} \\ 0 & \text{otherwise} \end{cases},$$

⁶The computation is performed using Julia 1.12.4.

$$y_{i1} = \begin{cases} 1 & \text{if } q(\eta_i, TD_1) > \zeta_{i1} \\ 0 & \text{otherwise} \end{cases},$$

$$p(\eta_i, y_{i,t-1}, TD_t) = \exp(\eta_i + \gamma y_{i,t-1} + TD_t) / (1 + \exp(\eta_i + \gamma y_{i,t-1} + TD_t)),$$

$$q(\eta_i, TD_1) = \exp(\eta_i + TD_1) / (1 + \exp(\eta_i + TD_1)),$$

$$\zeta_{it} \sim U(0, 1); \eta_i \sim N(0, \sigma_\eta^2).$$

The experiments are carried out with $\gamma = 1$, $TD_1 = 0.1$, $TD_2 = -0.1$, $TD_3 = 0.3$, $TD_4 = -0.3$, $TD_5 = -0.1$, $TD_6 = 0.3$, $TD_7 = 0.5$, $TD_8 = 0.2$, and $\sigma_\eta^2 = 0.5$. Accordingly, true values of the original parameters to be estimated (labelled true in the tables) are $\gamma = 1$, $\Delta TD_6 = 0.4$, $\Delta TD_7 = 0.2$, and $\Delta TD_8 = -0.3$.

In the experiments, the moment conditions (1) with $E[(1 \ y_{i,t-3})' \kappa_{it}] = 0$ excluded and (2) with $E[(1 \ y_{i,t-3})' \lambda_{it}] = 0$ excluded are used for just-identifying two distinct sets of the six transformed parameters, respectively. That is, we use both of the A estimator not using $E[(1 \ y_{i,t-3})' \kappa_{it}] = 0$ and B estimator not using $E[(1 \ y_{i,t-3})' \lambda_{it}] = 0$. The former is referred to as the A -(3+7) estimator, while the latter the B -(1+5) estimator.⁷

Using the results from these two estimators, two distinct sets of estimation results for the original parameters and their variances are obtained from the relationships among $\alpha_a^{(A)}$, $\alpha_b^{(A)}$, and $\alpha_d^{(A)}$ and among $\alpha_a^{(B)}$, $\alpha_b^{(B)}$, and $\alpha_d^{(B)}$, respectively (see details in appendices D and E). These Monte Carlo results for the original parameters are shown in Tables 1 and 2, respectively.⁸

Looking at the tables, all values of sd and rmse and almost all bias sizes decrease with the larger N , while values of se predict well those of sd. A calculation indicates that convergence rates of the estimators are root- N . Further, histograms and Q-Q plots generated from the estimates obtained using replication datasets depict asymptotic normality of the estimators.⁹

3.2 Model with time trends

In this case, $p(\eta_i, y_{i,t-1}, TD_t)$ and $q(\eta_i, TD_1)$ in section 3.1 are replaced by $p(\eta_i, y_{i,t-1}, t) = \exp(\eta_i + \gamma y_{i,t-1} + \varphi(t - \tau)) / (1 + \exp(\eta_i + \gamma y_{i,t-1} + \varphi(t - \tau)))$ and $q(\eta_i) = \exp(\eta_i + \varphi(1 -$

⁷Computational results using Maxima (2025) reveal that both of the sets of the moment functions composing the A -(3+7) and B -(1+5) estimators are linearly independent except where $(\gamma, \Delta TD_{t+1}) = (0, 0)$. In addition, both of the sets of the moment functions composing the A -(1+5) and B -(3+7) estimators are linearly independent except where $(\gamma, \Delta TD_{t+1}) = (0, 0)$ as well, where the former and the latter use the moment conditions (1) with $E[(1 \ y_{i,t-3})' \iota_{it}] = 0$ excluded and (2) with $E[(1 \ y_{i,t-3})' \mu_{it}] = 0$ excluded respectively. Accordingly, these four sets of the moment functions are also linearly independent except where $(\gamma, \varphi) = (0, 0)$ for the case of the model with time trends. Consequently, we clarify that using these four estimators, the models with time dummies and with time trends are point-identified except for the sets of true parameter values $(\gamma, \Delta TD_{t+1}) = (0, 0)$ and $(\gamma, \varphi) = (0, 0)$ respectively, as long as the rank conditions are satisfied (see also appendix C). Because of these facts, the usage of these four estimators is encouraged for root- N consistent estimations of the models with time dummies and with time trends.

⁸The corresponding results for the transformed parameters are shown in Tables F.1 and F.2 in appendix F, respectively.

⁹The figures are available from the author upon request.

Table 1: Monte Carlo results from $A-(3+7)$ estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
γ	1	0.99932	0.02739	0.02731	-0.00068	0.02740
ΔTD_6	0.4	0.39998	0.02044	0.02028	-0.00002	0.02044
ΔTD_7	0.2	0.19891	0.02760	0.02787	-0.00109	0.02762
ΔTD_8	-0.3	-0.29674	0.13188	0.13267	0.00326	0.13192
$N = 50$ million						
γ	1	0.99979	0.01203	0.01216	-0.00021	0.01203
ΔTD_6	0.4	0.40016	0.00889	0.00901	0.00016	0.00889
ΔTD_7	0.2	0.20042	0.01236	0.01235	0.00042	0.01237
ΔTD_8	-0.3	-0.29701	0.05900	0.05831	0.00299	0.05907
$N = 100$ million						
γ	1	0.99977	0.00878	0.00859	-0.00023	0.00878
ΔTD_6	0.4	0.40013	0.00651	0.00637	0.00013	0.00651
ΔTD_7	0.2	0.19990	0.00868	0.00873	-0.00010	0.00868
ΔTD_8	-0.3	-0.30004	0.04098	0.04104	-0.00004	0.04098

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation. 2) In each replication, ΔTD_6 is estimated with the two-step estimation and its standard error is calculated with the corrected variance.

Table 2: Monte Carlo results from $B-(1+5)$ estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
γ	1	1.00096	0.03788	0.03797	0.00096	0.03789
ΔTD_6	0.4	0.40105	0.02046	0.02029	0.00105	0.02048
ΔTD_7	0.2	0.19864	0.02851	0.02869	-0.00136	0.02854
ΔTD_8	-0.3	-0.33476	0.20280	0.20260	-0.03476	0.20576
$N = 50$ million						
γ	1	1.00036	0.01676	0.01677	0.00036	0.01676
ΔTD_6	0.4	0.40048	0.00898	0.00903	0.00048	0.00899
ΔTD_7	0.2	0.20039	0.01264	0.01268	0.00039	0.01265
ΔTD_8	-0.3	-0.30293	0.08340	0.08298	-0.00293	0.08345
$N = 100$ million						
γ	1	1.00028	0.01181	0.01185	0.00028	0.01182
ΔTD_6	0.4	0.40019	0.00640	0.00637	0.00019	0.00640
ΔTD_7	0.2	0.19988	0.00893	0.00896	-0.00012	0.00893
ΔTD_8	-0.3	-0.30359	0.05856	0.05853	-0.00359	0.05867

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation. 2) In each replication, ΔTD_6 is estimated with the two-step estimation and its standard error is calculated with the corrected variance.

$\tau)) / (1 + \exp(\eta_i + \varphi(1 - \tau)))$, respectively.

The experiments are carried out with $\gamma = 1$, $\varphi = 0.3$, $\tau = 1$, and $\sigma_\eta^2 = 0.5$; the former two are true values of the original parameters to be estimated.

In the experiments, using the results from the C estimator, the original parameters and their variances are estimated from the relationships among $\alpha_a^{(C)}$ and $\alpha_e^{(C)}$ (see details in appendix D). The Monte Carlo results are shown in Table 3.¹⁰

As well as for the case of previous subsection, all values of sd and rmse and all bias sizes decrease with the larger N , while values of se predict well those of sd. Root- N rate convergence and asymptotic normality of the estimator are confirmed as well as in previous subsection.

Table 3: Monte Carlo results from C estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
γ	1	1.00275	0.15319	0.15475	0.00275	0.15321
φ	0.3	0.29717	0.12971	0.13110	-0.00283	0.12974
$N = 50$ million						
γ	1	1.00038	0.06850	0.06876	0.00038	0.06851
φ	0.3	0.29951	0.05788	0.05819	-0.00049	0.05788
$N = 100$ million						
γ	1	0.99994	0.05019	0.04860	-0.00006	0.05019
φ	0.3	0.29997	0.04250	0.04113	-0.00003	0.04250

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation.

4 DISCUSSION AND CONCLUSION

Honoré and Tamer (2006) worry that no point-identification will be conducted for dynamic fixed effects binary choice models only with time effects, scrutinizing their results obtained with a linear programming method under probit specifications.¹¹ However, other than for special true values of parameters of interest and atypical datasets, this worry turns out to be needless as long as five consecutive time periods are provided and logit specifications are used, as seen from the results in previous sections.

As for the model with time dummies, in Figures 5 and 6 (where four consecutive time periods are used) in Honoré and Tamer (2006), horizontal sizes of the upper three identified

¹⁰The corresponding results for the transformed parameters are shown in Table F.3 in appendix F.

¹¹Since the probit and logit functions are similarly shaped, it is reasonably conceivable that the results from using logit specifications will be almost the same as those from using probit ones.

regions are vanishingly small. In addition, the identified regions for (γ, δ_3) and (γ, δ_4) are congruent to that for (γ, δ_2) ; the former two are vertically lowered from the latter one by the sizes of $\delta_2 - \delta_3$ and $\delta_2 - \delta_4$ (i.e., 0.1 and 0.2), respectively. Further, the lower identified regions for (δ_3, δ_2) , (δ_4, δ_2) and (δ_4, δ_3) are almost the 45-degree line segments with the intercepts being the values of $\delta_3 - \delta_2$, $\delta_4 - \delta_2$ and $\delta_4 - \delta_3$ (i.e., 0.1, 0.2 and 0.1), respectively. Therefore, identified regions for $(\gamma, \delta_3 - \delta_2, \delta_4 - \delta_3)$ are almost singleton.¹² Further, Figures 4 (where three consecutive time periods are used) and 5 for $\gamma = 1$ are similar in shape, although sizable identified regions in the former cannot in any way be regarded as singleton.

As for the model with time trends, all nine identified regions for (γ, β) in Figure 3 (where four consecutive time periods are used) in Honoré and Tamer (2006) are vanishingly small and almost singleton, although the identified regions in Figure 2 (where three consecutive time periods are used) are of non-negligible size.¹³

It might be inferred from the foregoing that the identified regions obtained from the linear programming method will accurately reduce to singleton for both dynamic fixed effects logit models only with time dummies and only with time trends if five consecutive time periods are provided. This inference dovetails with the theoretical and experimental results in this paper.

The root- N consistent estimators proposed in this paper settled the problem whether dynamic fixed effects logit models only with time effects are point-identified or not. The discovery of root- N consistent estimators was in an open-and-shut manner materialized for dynamic fixed effects logit models only with time effects. Exactly, “the darkest place is under the candlestick.”

Postscript: To enhance the authenticity of the estimators proposed in this paper, we can conduct the Wald test of the restrictions among the transformed parameters using their estimates (see appendix G).

ACKNOWLEDGMENTS

The author would like to acknowledge the use of generative AI tools, specifically OpenAI ChatGPT and Google Gemini, for technical assistance in optimizing the Julia scripts used for Monte Carlo simulations and providing conceptual clarifications on mathematical methods. The author also extends the gratitude to the Stack Exchange community for providing valuable insights into programming challenges. In addition, the AI tools were utilized for linguistic refinement of the manuscript. All AI-generated suggestions and community-sourced solutions were critically reviewed and validated by the author.

¹²Note that γ , δ_2 , δ_3 , and δ_4 in Honoré and Tamer (2006) correspond to γ , TD_{t-1} , TD_t , and TD_{t+1} in section 2, respectively. Accordingly, $\delta_3 - \delta_2$ and $\delta_4 - \delta_3$ correspond to ΔTD_t and ΔTD_{t+1} , respectively.

¹³Note that γ and β in Honoré and Tamer (2006) correspond to γ and φ in section 2, respectively.

REFERENCES

- Bezanson, J., A. Edelman, S. Karpinski, and V.B. Shah (2017) “Julia: A fresh approach to numerical computing,” *SIAM review*, 59 (1), 65–98, 10.1137/141000671.
- Bonhomme, S. (2012) “Functional differencing,” *Econometrica*, 80 (4), 1337–1385, 10.3982/ECTA9311.
- Chamberlain, G. (1985) “Heterogeneity, omitted variable bias, and duration dependence,” in J.J. Heckman and B. Singer eds. *Longitudinal Analysis of Labor Market Data*, 3–38, Cambridge: Cambridge University Press, 10.1017/CCOL0521304539.001.
- Dano, K. (2023) “Transition probabilities and identifying moments in dynamic fixed effects logit models,” February, 10.48550/arXiv.2303.00083, arXiv:2303.00083 [econ.EM].
- “Kevin Dano’s Research Website,” <https://kevindano.github.io/research>, (Accessed: 2026-02-06).
- Dobronyi, C., J. Gu, and K. Kim (2021) “Identification of dynamic panel logit models with fixed effects,” April, 10.48550/arXiv.2104.04590, arXiv:2104.04590 [econ.EM] (Note: In the version dated January 3, 2025, T.M. Russell joins as one of the coauthors).
- Honoré, B.E. and E. Kyriazidou (2000) “Panel data discrete choice models with lagged dependent variables,” *Econometrica*, 68 (4), 839–874, 10.1111/1468-0262.00139.
- Honoré, B.E. and E. Tamer (2006) “Bounds on parameters in panel dynamic discrete choice models,” *Econometrica*, 74 (3), 611–629, 10.1111/j.1468-0262.2006.00676.x.
- Honoré, B.E. and M. Weidner (2025) “Moment conditions for dynamic panel logit models with fixed effects,” *Review of Economic Studies*, 92 (5), 3112–3137, 10.1093/restud/rdae097.
- Kitazawa, Y. (2022) “Transformations and moment conditions for dynamic fixed effects logit models,” *Journal of Econometrics*, 229 (2), 350–362, 10.1016/j.jeconom.2021.01.007.
- Kruiniger, H. (2020) “Further results on the estimation of dynamic panel logit models with fixed effects,” October, 10.48550/arXiv.2010.03382, arXiv:2010.03382 [econ.EM].
- Maxima (2025) “Maxima, a Computer Algebra System. Version 5.48.1,” <https://maxima.sourceforge.io/>.
- Murphy, K.M. and R.H. Topel (1985) “Estimation and inference in two-step econometric models,” *Journal of Business & Economic Statistics*, 3 (4), 370–379, 10.1080/07350015.1985.10509471.
- Newey, W.K. (1984) “A method of moments interpretation of sequential estimators,” *Economics Letters*, 14 (2), 201–206, 10.1016/0165-1765(84)90083-1.

Pagan, A. (1984) “Econometric issues in the analysis of regressions with generated regressors,” *International Economic Review*, 25 (1), 221–247, 10.2307/2648877.

Ruud, P.A. (2000) *An Introduction to Classical Econometric Theory*, New York: Oxford University Press.

Taboga, M. (2021) “Delta method,” <https://www.statlect.com/asymptotic-theory/delta-method>, Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. (Accessed: 2022-06-21).

APPENDICES

A Derivation of linear transformations

We derive the linear transformations of moment conditions based on the *g-form* and *h-form*.

First, given that $(1 - y_{i,t-2})y_{i,t-2} = 0$,

$$\begin{aligned}
& (\phi_t \phi_{t+1} + 1) \phi_{t+1} (1 - y_{i,t-2}) \hbar U_{it}^- \\
&= (\phi_t \phi_{t+1} + 1) (1 - y_{i,t-2}) (U_{it}^- - y_{i,t-1}) \phi_{t+1} \\
&\quad + (\phi_t \phi_{t+1} - 1) (1 - y_{i,t-2}) ((U_{it}^- - y_{i,t-1}) - 2U_{it}^- (1 - y_{i,t-1})) \phi_{t+1} \\
&= (\phi_t \phi_{t+1} + 1) (1 - y_{i,t-2}) (U_{it}^- (1 - y_{i,t-1}) + y_{i,t-1} (U_{it}^- - y_{i,t-1})) \phi_{t+1} \\
&\quad + (\phi_t \phi_{t+1} - 1) (1 - y_{i,t-2}) (U_{it}^- (1 - y_{i,t-1}) + y_{i,t-1} (U_{it}^- - y_{i,t-1})) \phi_{t+1} \\
&\quad - 2(\phi_t \phi_{t+1} - 1) (1 - y_{i,t-2}) U_{it}^- (1 - y_{i,t-1}) \phi_{t+1} \\
&= 2(1 - y_{i,t-2}) U_{it}^- (1 - y_{i,t-1}) \phi_{t+1} \\
&\quad + 2\phi_t \phi_{t+1} (1 - y_{i,t-2}) y_{i,t-1} (U_{it}^- - y_{i,t-1}) \phi_{t+1} \\
&= 2(1 - y_{i,t-2}) (1 - y_{i,t-1}) ((y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} - (1 - y_{it}) y_{i,t+1}) \\
&\quad + 2\phi_t \phi_{t+1} (1 - y_{i,t-2}) y_{i,t-1} \\
&\quad \times (((y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} - (1 - y_{it}) y_{i,t+1} - \delta y_{i,t-1} (1 - y_{it}) y_{i,t+1}) \\
&\quad \quad - y_{i,t-1} \phi_{t+1}) \\
&= 2(1 - y_{i,t-2}) (1 - y_{i,t-1}) (y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} \\
&\quad - 2(1 - y_{i,t-2}) (1 - y_{i,t-1}) (1 - y_{it}) y_{i,t+1} \\
&\quad + 2(1 - y_{i,t-2}) y_{i,t-1} (y_{it} + (1 - y_{it}) y_{i,t+1} - y_{i,t-1}) \phi_t (\phi_{t+1})^2 \\
&\quad - 2(1 - y_{i,t-2}) y_{i,t-1} (1 - y_{it}) y_{i,t+1} \phi_t \phi_{t+1} (1 + \delta),
\end{aligned}$$

where the relationships $(1 - y_{i,t-1})y_{i,t-1} = 0$ and $y_{i,t-1}^2 = y_{i,t-1}$ are also used. Multiplying $(\phi_t \phi_{t+1} + 1) \phi_{t+1} (1 - y_{i,t-2}) \hbar U_{it}^-$ by $(1/2) \phi_{t+1}^{-1}$ and $(1/2) \phi_t^{-1} \phi_{t+1}^{-1} (1 + \delta)^{-1}$ gives ι_{it} and λ_{it} , respectively.

Then, given that $y_{i,t-2}^2 = y_{i,t-2}$,

$$(\phi_t \phi_{t+1} + (\delta + 1)) \phi_{t+1} y_{i,t-2} \hbar U_{it}^-$$

$$\begin{aligned}
&= (\phi_t \phi_{t+1} + (\delta + 1)) y_{i,t-2} (U_{it}^- - y_{i,t-1}) \phi_{t+1} \\
&\quad + (\phi_t \phi_{t+1} - (\delta + 1)) y_{i,t-2} ((U_{it}^- - y_{i,t-1}) - 2U_{it}^- (1 - y_{i,t-1})) \phi_{t+1} \\
&= (\phi_t \phi_{t+1} + (\delta + 1)) y_{i,t-2} (U_{it}^- (1 - y_{i,t-1}) + y_{i,t-1} (U_{it}^- - y_{i,t-1})) \phi_{t+1} \\
&\quad + (\phi_t \phi_{t+1} - (\delta + 1)) y_{i,t-2} (U_{it}^- (1 - y_{i,t-1}) + y_{i,t-1} (U_{it}^- - y_{i,t-1})) \phi_{t+1} \\
&\quad - 2(\phi_t \phi_{t+1} - (\delta + 1)) y_{i,t-2} U_{it}^- (1 - y_{i,t-1}) \phi_{t+1} \\
&= 2(\delta + 1) y_{i,t-2} U_{it}^- (1 - y_{i,t-1}) \phi_{t+1} \\
&\quad + 2\phi_t \phi_{t+1} y_{i,t-2} y_{i,t-1} (U_{it}^- - y_{i,t-1}) \phi_{t+1} \\
&= 2(\delta + 1) y_{i,t-2} (1 - y_{i,t-1}) ((y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} - (1 - y_{it}) y_{i,t+1}) \\
&\quad + 2\phi_t \phi_{t+1} y_{i,t-2} y_{i,t-1} \\
&\quad \times ((y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} - (1 - y_{it}) y_{i,t+1} - \delta y_{i,t-1} (1 - y_{it}) y_{i,t+1}) \\
&\quad \quad - y_{i,t-1} \phi_{t+1}) \\
&= 2y_{i,t-2} (1 - y_{i,t-1}) (y_{it} + (1 - y_{it}) y_{i,t+1}) \phi_{t+1} (1 + \delta) \\
&\quad - 2y_{i,t-2} (1 - y_{i,t-1}) (1 - y_{it}) y_{i,t+1} (1 + \delta) \\
&\quad + 2y_{i,t-2} y_{i,t-1} (y_{it} + (1 - y_{it}) y_{i,t+1} - y_{i,t-1}) \phi_t (\phi_{t+1})^2 \\
&\quad - 2y_{i,t-2} y_{i,t-1} (1 - y_{it}) y_{i,t+1} \phi_t \phi_{t+1} (1 + \delta),
\end{aligned}$$

where the relationships $(1 - y_{i,t-1}) y_{i,t-1} = 0$ and $y_{i,t-1}^2 = y_{i,t-1}$ are also used. Multiplying $(\phi_t \phi_{t+1} + (\delta + 1)) \phi_{t+1} y_{i,t-2} \hbar U_{it}^-$ by $(1/2) \phi_{t+1}^{-1} (1 + \delta)^{-1}$ and $(1/2) \phi_t^{-1} \phi_{t+1}^{-1} (1 + \delta)^{-1}$ gives ι_{it}^* and λ_{it}^* , respectively.

Then, given that $y_{i,t-2} (1 - y_{i,t-2}) = 0$,

$$\begin{aligned}
&(\phi_t^{-1} \phi_{t+1}^{-1} + 1) \phi_{t+1}^{-1} y_{i,t-2} \hbar \Upsilon_{it}^- \\
&= (\phi_t^{-1} \phi_{t+1}^{-1} + 1) y_{i,t-2} (\Upsilon_{it}^- - y_{i,t-1}) \phi_{t+1}^{-1} \\
&\quad + (\phi_t^{-1} \phi_{t+1}^{-1} - 1) y_{i,t-2} ((\Upsilon_{it}^- - y_{i,t-1}) - 2(\Upsilon_{it}^- - y_{i,t-1}) y_{i,t-1}) \phi_{t+1}^{-1} \\
&= (\phi_t^{-1} \phi_{t+1}^{-1} + 1) y_{i,t-2} ((\Upsilon_{it}^- - y_{i,t-1}) y_{i,t-1} + (1 - y_{i,t-1}) \Upsilon_{it}^-) \phi_{t+1}^{-1} \\
&\quad + (\phi_t^{-1} \phi_{t+1}^{-1} - 1) y_{i,t-2} ((\Upsilon_{it}^- - y_{i,t-1}) y_{i,t-1} + (1 - y_{i,t-1}) \Upsilon_{it}^-) \phi_{t+1}^{-1} \\
&\quad - 2(\phi_t^{-1} \phi_{t+1}^{-1} - 1) y_{i,t-2} (\Upsilon_{it}^- - y_{i,t-1}) y_{i,t-1} \phi_{t+1}^{-1} \\
&= 2y_{i,t-2} (\Upsilon_{it}^- - y_{i,t-1}) y_{i,t-1} \phi_{t+1}^{-1} \\
&\quad + 2\phi_t^{-1} \phi_{t+1}^{-1} y_{i,t-2} (1 - y_{i,t-1}) \Upsilon_{it}^- \phi_{t+1}^{-1} \\
&= 2y_{i,t-2} y_{i,t-1} (y_{it} y_{i,t+1} \phi_{t+1}^{-1} + y_{it} (1 - y_{i,t+1}) - y_{t-1} \phi_{t+1}^{-1}) \\
&\quad + 2\phi_t^{-1} \phi_{t+1}^{-1} y_{i,t-2} (1 - y_{i,t-1}) \\
&\quad \quad \times (y_{it} y_{i,t+1} \phi_{t+1}^{-1} + y_{it} (1 - y_{i,t+1}) + \delta (1 - y_{i,t-1}) y_{it} (1 - y_{i,t+1})) \\
&= 2y_{i,t-2} y_{i,t-1} (y_{it} y_{i,t+1} - y_{i,t-1}) \phi_{t+1}^{-1} \\
&\quad + 2y_{i,t-2} y_{i,t-1} y_{it} (1 - y_{i,t+1}) \\
&\quad + 2y_{i,t-2} (1 - y_{i,t-1}) y_{it} y_{i,t+1} \phi_t^{-1} (\phi_{t+1}^{-1})^2 \\
&\quad + 2y_{i,t-2} (1 - y_{i,t-1}) y_{it} (1 - y_{i,t+1}) \phi_t^{-1} \phi_{t+1}^{-1} (1 + \delta),
\end{aligned}$$

where the relationships $y_{i,t-1} (1 - y_{i,t-1}) = 0$ and $(1 - y_{i,t-1})^2 = 1 - y_{i,t-1}$ are also used.

Multiplying $(\phi_t^{-1}\phi_{t+1}^{-1} + 1)\phi_{t+1}^{-1}y_{i,t-2}\hbar\Upsilon_{it}^-$ by $(1/2)\phi_t\phi_{t+1}(1+\delta)^{-1}$ and $(1/2)\phi_{t+1}$ gives κ_{it} and μ_{it} , respectively.

Then, given that $(1 - y_{i,t-2})^2 = (1 - y_{i,t-2})$,

$$\begin{aligned}
& (\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))\phi_{t+1}^{-1}(1 - y_{i,t-2})\hbar\Upsilon_{it}^- \\
&= (\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))(1 - y_{i,t-2})(\Upsilon_{it}^- - y_{i,t-1})\phi_{t+1}^{-1} \\
&\quad + (\phi_t^{-1}\phi_{t+1}^{-1} - (\delta + 1))(1 - y_{i,t-2})((\Upsilon_{it}^- - y_{i,t-1}) - 2(\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1})\phi_{t+1}^{-1} \\
&= (\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))(1 - y_{i,t-2})((\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1} + (1 - y_{i,t-1})\Upsilon_{it}^-)\phi_{t+1}^{-1} \\
&\quad + (\phi_t^{-1}\phi_{t+1}^{-1} - (\delta + 1))(1 - y_{i,t-2})((\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1} + (1 - y_{i,t-1})\Upsilon_{it}^-)\phi_{t+1}^{-1} \\
&\quad - 2(\phi_t^{-1}\phi_{t+1}^{-1} - (\delta + 1))(1 - y_{i,t-2})(\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1}\phi_{t+1}^{-1} \\
&= 2(\delta + 1)(1 - y_{i,t-2})(\Upsilon_{it}^- - y_{i,t-1})y_{i,t-1}\phi_{t+1}^{-1} \\
&\quad + 2\phi_t^{-1}\phi_{t+1}^{-1}(1 - y_{i,t-2})(1 - y_{i,t-1})\Upsilon_{it}^-\phi_{t+1}^{-1} \\
&= 2(\delta + 1)(1 - y_{i,t-2})y_{i,t-1}(y_{it}y_{i,t+1}\phi_{t+1}^{-1} + y_{it}(1 - y_{i,t+1}) - y_{t-1}\phi_{t+1}^{-1}) \\
&\quad + 2\phi_t^{-1}\phi_{t+1}^{-1}(1 - y_{i,t-2})(1 - y_{i,t-1}) \\
&\quad \quad \times (y_{it}y_{i,t+1}\phi_{t+1}^{-1} + y_{it}(1 - y_{i,t+1}) + \delta(1 - y_{i,t-1})y_{it}(1 - y_{i,t+1})) \\
&= 2(1 - y_{i,t-2})y_{i,t-1}(y_{it}y_{i,t+1} - y_{i,t-1})\phi_{t+1}^{-1}(1 + \delta) \\
&\quad + 2(1 - y_{i,t-2})y_{i,t-1}y_{it}(1 - y_{i,t+1})(1 + \delta) \\
&\quad + 2(1 - y_{i,t-2})(1 - y_{i,t-1})y_{it}y_{i,t+1}\phi_t^{-1}(\phi_{t+1}^{-1})^2 \\
&\quad + 2(1 - y_{i,t-2})(1 - y_{i,t-1})y_{it}(1 - y_{i,t+1})\phi_t^{-1}\phi_{t+1}^{-1}(1 + \delta),
\end{aligned}$$

where the relationships $y_{i,t-1}(1 - y_{i,t-1}) = 0$ and $(1 - y_{i,t-1})^2 = 1 - y_{i,t-1}$ are also used. Multiplying $(\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))\phi_{t+1}^{-1}(1 - y_{i,t-2})\hbar\Upsilon_{it}^-$ by $(1/2)\phi_t\phi_{t+1}(1+\delta)^{-1}$ and $(1/2)\phi_{t+1}(1+\delta)^{-1}$ gives κ_{it}^* and μ_{it}^* , respectively.

Finally, with $\phi = \phi_t = \phi_{t+1}$, $(\phi_t\phi_{t+1} + 1)\phi_{t+1}(1 - y_{i,t-2})\hbar U_{it}^-$, $(\phi_t\phi_{t+1} + (\delta + 1))\phi_{t+1}y_{i,t-2}\hbar U_{it}^-$, $(\phi_t^{-1}\phi_{t+1}^{-1} + 1)\phi_{t+1}^{-1}y_{i,t-2}\hbar\Upsilon_{it}^-$, and $(\phi_t^{-1}\phi_{t+1}^{-1} + (\delta + 1))\phi_{t+1}^{-1}(1 - y_{i,t-2})\hbar\Upsilon_{it}^-$ are as follows:

$$\begin{aligned}
& (\phi^2 + 1)\phi(1 - y_{i,t-2})\hbar U_{it}^- \\
&= 2(1 - y_{i,t-2})(1 - y_{i,t-1})(y_{it} + (1 - y_{it})y_{i,t+1})\phi \\
&\quad - 2(1 - y_{i,t-2})(1 - y_{i,t-1})(1 - y_{it})y_{i,t+1} \\
&\quad + 2(1 - y_{i,t-2})y_{i,t-1}(y_{it} + (1 - y_{it})y_{i,t+1} - y_{i,t-1})\phi^3 \\
&\quad - 2(1 - y_{i,t-2})y_{i,t-1}(1 - y_{it})y_{i,t+1}\phi^2(1 + \delta),
\end{aligned}$$

$$\begin{aligned}
& (\phi^2 + (\delta + 1))\phi y_{i,t-2}\hbar U_{it}^- \\
&= 2y_{i,t-2}(1 - y_{i,t-1})(y_{it} + (1 - y_{it})y_{i,t+1})\phi(1 + \delta) \\
&\quad - 2y_{i,t-2}(1 - y_{i,t-1})(1 - y_{it})y_{i,t+1}(1 + \delta) \\
&\quad + 2y_{i,t-2}y_{i,t-1}(y_{it} + (1 - y_{it})y_{i,t+1} - y_{i,t-1})\phi^3 \\
&\quad - 2y_{i,t-2}y_{i,t-1}(1 - y_{it})y_{i,t+1}\phi^2(1 + \delta),
\end{aligned}$$

$$(\phi^{-2} + 1)\phi^{-1}y_{i,t-2}\hbar\Upsilon_{it}^-$$

$$\begin{aligned}
&= 2y_{i,t-2}y_{i,t-1}(y_{it}y_{i,t+1} - y_{i,t-1})\phi^{-1} \\
&\quad + 2y_{i,t-2}y_{i,t-1}y_{it}(1 - y_{i,t+1}) \\
&\quad + 2y_{i,t-2}(1 - y_{i,t-1})y_{it}y_{i,t+1}\phi^{-3} \\
&\quad + 2y_{i,t-2}(1 - y_{i,t-1})y_{it}(1 - y_{i,t+1})\phi^{-2}(1 + \delta),
\end{aligned}$$

$$\begin{aligned}
&(\phi^{-2} + (\delta + 1))\phi^{-1}(1 - y_{i,t-2})\hbar\Upsilon_{it}^- \\
&= 2(1 - y_{i,t-2})y_{i,t-1}(y_{it}y_{i,t+1} - y_{i,t-1})\phi^{-1}(1 + \delta) \\
&\quad + 2(1 - y_{i,t-2})y_{i,t-1}y_{it}(1 - y_{i,t+1})(1 + \delta) \\
&\quad + 2(1 - y_{i,t-2})(1 - y_{i,t-1})y_{it}y_{i,t+1}\phi^{-3} \\
&\quad + 2(1 - y_{i,t-2})(1 - y_{i,t-1})y_{it}(1 - y_{i,t+1})\phi^{-2}(1 + \delta).
\end{aligned}$$

Multiplying $(\phi^2 + 1)\phi(1 - y_{i,t-2})\hbar U_{it}^-$ by $(1/2)\phi^{-1}$, $(\phi^2 + (\delta + 1))\phi y_{i,t-2}\hbar U_{it}^-$ by $(1/2)\phi^{-2}(1 + \delta)^{-1}$, $(\phi^{-2} + 1)\phi^{-1}y_{i,t-2}\hbar\Upsilon_{it}^-$ by $(1/2)\phi$, and $(\phi^{-2} + (\delta + 1))\phi^{-1}(1 - y_{i,t-2})\hbar\Upsilon_{it}^-$ by $(1/2)\phi^2(1 + \delta)^{-1}$ gives ϱ_{it} , ϱ_{it}^* , ς_{it} , and ς_{it}^* , respectively.

B Estimators for transformed parameters

To begin with, we define

$$\bar{\Theta}_t^{(1-)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(1)} (1 - y_{i,t-2}),$$

$$\bar{\Theta}_t^{(2-)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(2)} (1 - y_{i,t-2}),$$

$$\bar{\Theta}_t^{(3-)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(3)} (1 - y_{i,t-2}),$$

$$\bar{\Theta}_t^{(4-)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(4)} (1 - y_{i,t-2}),$$

$$\bar{\Theta}_t^{(1+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(1)} y_{i,t-2},$$

$$\bar{\Theta}_t^{(2+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(2)} y_{i,t-2},$$

$$\bar{\Theta}_t^{(3+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(3)} y_{i,t-2},$$

$$\bar{\Theta}_t^{(4+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(4)} y_{i,t-2},$$

$$\bar{\Xi}_t^{(1+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(1)} y_{i,t-2},$$

$$\bar{\Xi}_t^{(2+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(2)} y_{i,t-2},$$

$$\bar{\Xi}_t^{(3+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(3)} y_{i,t-2},$$

$$\bar{\Xi}_t^{(4+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(4)} y_{i,t-2},$$

$$\bar{\Xi}_t^{(1-)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(1)} (1 - y_{i,t-2}),$$

$$\bar{\Xi}_t^{(2-)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(2)} (1 - y_{i,t-2}),$$

$$\bar{\Xi}_t^{(3-)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(3)} (1 - y_{i,t-2}),$$

$$\bar{\Xi}_t^{(4-)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(4)} (1 - y_{i,t-2}),$$

$$\bar{\Theta}_t^{(1-+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(1)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Theta}_t^{(2-+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(2)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Theta}_t^{(3-+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(3)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Theta}_t^{(4-+)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(4)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Theta}_t^{(1++)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(1)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Theta}_t^{(2++)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(2)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Theta}_t^{(3++)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(3)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Theta}_t^{(4++)} = \frac{1}{N} \sum_{i=1}^N \Theta_{it}^{(4)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Xi}_t^{(1++)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(1)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Xi}_t^{(2++)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(2)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Xi}_t^{(3++)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(3)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Xi}_t^{(4++)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(4)} y_{i,t-2} y_{i,t-3},$$

$$\bar{\Xi}_t^{(1-+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(1)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Xi}_t^{(2-+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(2)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Xi}_t^{(3-+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(3)} (1 - y_{i,t-2}) y_{i,t-3},$$

$$\bar{\Xi}_t^{(4-+)} = \frac{1}{N} \sum_{i=1}^N \Xi_{it}^{(4)} (1 - y_{i,t-2}) y_{i,t-3}.$$

First, we write the sample analogs of the moment conditions (1) in a vector-matrix form to derive the A estimator and its estimated asymptotic variance-covariance matrix.

Putting

$$\bar{Y}_t^{(A)} = \begin{bmatrix} -\bar{\Theta}_t^{(1-)} \\ -\bar{\Theta}_t^{(1+)} \\ -\bar{\Xi}_t^{(4+)} \\ -\bar{\Xi}_t^{(4-)} \\ -\bar{\Theta}_t^{(1-+)} \\ -\bar{\Theta}_t^{(1++)} \\ -\bar{\Xi}_t^{(4++)} \\ -\bar{\Xi}_t^{(4-+)} \end{bmatrix}, \quad \theta^{(A)} = \begin{bmatrix} \alpha_a^{(A)} \\ \alpha_b^{(A)} \\ \alpha_c^{(A)} \\ \alpha_d^{(A)} \\ \alpha_e^{(A)} \\ \alpha_f^{(A)} \\ \alpha_g^{(A)} \end{bmatrix},$$

and

$$\bar{X}_t^{(A)} = \begin{bmatrix} 0 & \bar{\Theta}_t^{(2-)} & \bar{\Theta}_t^{(3-)} & \bar{\Theta}_t^{(4-)} & 0 & 0 & 0 \\ \bar{\Theta}_t^{(4+)} & \bar{\Theta}_t^{(2+)} & 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(3+)} \\ 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(1+)} & \bar{\Xi}_t^{(3+)} & \bar{\Xi}_t^{(2+)} \\ \bar{\Xi}_t^{(1-)} & 0 & \bar{\Xi}_t^{(2-)} & 0 & 0 & \bar{\Xi}_t^{(3-)} & 0 \\ 0 & \bar{\Theta}_t^{(2-+)} & \bar{\Theta}_t^{(3-+)} & \bar{\Theta}_t^{(4-+)} & 0 & 0 & 0 \\ \bar{\Theta}_t^{(4++)} & \bar{\Theta}_t^{(2++)} & 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(3++)} \\ 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(1++)} & \bar{\Xi}_t^{(3++)} & \bar{\Xi}_t^{(2++)} \\ \bar{\Xi}_t^{(1-+)} & 0 & \bar{\Xi}_t^{(2-+)} & 0 & 0 & \bar{\Xi}_t^{(3-+)} & 0 \end{bmatrix},$$

we write the sample analog as follows:

$$\bar{Y}_t^{(A)} = \bar{X}_t^{(A)} \theta^{(A)}.$$

Using $\bar{Y}_t^{(A-r)}$ and $\bar{X}_t^{(A-r)}$, both of which remove the r th row from $\bar{Y}_t^{(A)}$ and $\bar{X}_t^{(A)}$ respectively, we write the sample analog as follows:

$$\bar{Y}_t^{(A-r)} = \bar{X}_t^{(A-r)} \theta^{(A)}.$$

The A estimator not using the r th moment condition is solved as follows:

$$\hat{\theta}^{(A[-r])} = \left(\bar{X}_t^{(A-r)} \right)^{-1} \bar{Y}_t^{(A-r)},$$

as long as $\bar{X}_t^{(A-r)}$ is invertible.

The residual vector of this A estimator is as follows:

$$\hat{V}_{it}^{(A-r)} = Y_{it}^{(A-r)} - X_{it}^{(A-r)} \hat{\theta}^{(A[-r])},$$

where $Y_{it}^{(A-r)}$ and $X_{it}^{(A-r)}$ are the i th summand in the calculation of the averages $\bar{Y}_t^{(A-r)}$ and $\bar{X}_t^{(A-r)}$, respectively.

Defining

$$\hat{W}_t^{(A-r)} = \left((1/N) \sum_{i=1}^N \hat{V}_{it}^{(A-r)} (\hat{V}_{it}^{(A-r)})' \right)^{-1},$$

the estimated asymptotic variance-covariance matrix of this A estimator is

$$\widehat{\text{Var}}(\hat{\theta}^{(A[-r])}) = (1/N) \left((\bar{X}_t^{(A-r)})' \hat{W}_t^{(A-r)} \bar{X}_t^{(A-r)} \right)^{-1}.$$

Moreover, the A estimator not using the moment conditions $E[(1 \ y_{i,t-3})' \kappa_{it}] = 0$ is defined by replacing r in the superscripts appended to the variables and statistics above by $(3+7)$. The expression $-(3+7)$ implies the removal of both the third and seventh rows from the vector $\bar{Y}_t^{(A)}$ and matrix $\bar{X}_t^{(A)}$. In this case, the fifth column of $\bar{X}_t^{(A-(3+7))}$

is also removed since all elements in this column are zero-valued, and accordingly instead of $\theta^{(A)}$, $\theta_{[-e]}^{(A)}$ is used, which is the parameter vector after removing $\alpha_e^{(A)}$ from $\theta^{(A)}$, and the estimator for it is written as $\hat{\theta}_{[-e]}^{(A[-(3+7)])}$. The subsequent statistics including the asymptotic variance-covariance matrix are constructed using them. Similarly, the A estimator not using $E[(1 \ y_{i,t-3})' \ \nu_{it}] = 0$ and its asymptotic variance-covariance matrix are defined by replacing r by $(1 + 5)$ and using $\hat{\theta}_{[-d]}^{(A[-(1+5)])}$ for $\theta_{[-d]}^{(A)}$.

Next, we describe the B estimator and its estimated asymptotic variance-covariance matrix, using the sample analogs of the moment conditions (2).

Putting

$$\bar{Y}_t^{(B)} = \begin{bmatrix} -\bar{\Theta}_t^{(4-)} \\ -\bar{\Theta}_t^{(4+)} \\ -\bar{\Xi}_t^{(1+)} \\ -\bar{\Xi}_t^{(1-)} \\ -\bar{\Theta}_t^{(4-+)} \\ -\bar{\Theta}_t^{(4++)} \\ -\bar{\Xi}_t^{(1++)} \\ -\bar{\Xi}_t^{(1-+)} \end{bmatrix}, \quad \theta^{(B)} = \begin{bmatrix} \alpha_a^{(B)} \\ \alpha_b^{(B)} \\ \alpha_c^{(B)} \\ \alpha_d^{(B)} \\ \alpha_e^{(B)} \\ \alpha_f^{(B)} \\ \alpha_g^{(B)} \end{bmatrix},$$

and

$$\bar{X}_t^{(B)} = \begin{bmatrix} 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(1-)} & \bar{\Theta}_t^{(3-)} & \bar{\Theta}_t^{(2-)} \\ \bar{\Theta}_t^{(1+)} & 0 & \bar{\Theta}_t^{(2+)} & 0 & 0 & \bar{\Theta}_t^{(3+)} & 0 \\ 0 & \bar{\Xi}_t^{(2+)} & \bar{\Xi}_t^{(3+)} & \bar{\Xi}_t^{(4+)} & 0 & 0 & 0 \\ \bar{\Xi}_t^{(4-)} & \bar{\Xi}_t^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-)} \\ 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(1-+)} & \bar{\Theta}_t^{(3-+)} & \bar{\Theta}_t^{(2-+)} \\ \bar{\Theta}_t^{(1++)} & 0 & \bar{\Theta}_t^{(2++)} & 0 & 0 & \bar{\Theta}_t^{(3++)} & 0 \\ 0 & \bar{\Xi}_t^{(2++)} & \bar{\Xi}_t^{(3++)} & \bar{\Xi}_t^{(4++)} & 0 & 0 & 0 \\ \bar{\Xi}_t^{(4-+)} & \bar{\Xi}_t^{(2-+)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-+)} \end{bmatrix},$$

we write the sample analog of the moment conditions (10) as follows:

$$\bar{Y}_t^{(B)} = \bar{X}_t^{(B)} \theta^{(B)}.$$

The B estimator not using the r th moment condition and its estimated asymptotic variance-covariance matrix are defined, replacing a symbol A by B in the superscripts appended to the variables and statistics regarding the A estimator. Moreover, the B estimator not using $E[(1 \ y_{i,t-3})' \ \lambda_{it}] = 0$ and its asymptotic variance-covariance matrix is defined by replacing r by $(1 + 5)$ and using $\hat{\theta}_{[-e]}^{(B[-(1+5)])}$ for $\theta_{[-e]}^{(B)}$, while the B estimator not using $E[(1 \ y_{i,t-3})' \ \mu_{it}] = 0$ and its asymptotic variance-covariance matrix is defined by replacing r by $(3 + 7)$ and using $\hat{\theta}_{[-d]}^{(B[-(3+7)])}$ for $\theta_{[-d]}^{(B)}$.

Third, we describe the C estimator and its estimated asymptotic variance-covariance matrix, using the sample analogs of the moment conditions (3).

Putting

$$\bar{Y}_t^{(C)} = \begin{bmatrix} -\bar{\Theta}_t^{(1-)} \\ -\bar{\Theta}_t^{(4+)} \\ -\bar{\Xi}_t^{(1+)} \\ -\bar{\Xi}_t^{(4-)} \\ -\bar{\Theta}_{t-1}^{(1-)} \\ -\bar{\Theta}_{t-1}^{(4+)} \\ -\bar{\Xi}_{t-1}^{(1+)} \\ -\bar{\Xi}_{t-1}^{(4-)} \end{bmatrix}, \quad \theta^{(C)} = \begin{bmatrix} \alpha_a^{(C)} \\ \alpha_b^{(C)} \\ \alpha_c^{(C)} \\ \alpha_d^{(C)} \\ \alpha_e^{(C)} \\ \alpha_f^{(C)} \\ \alpha_g^{(C)} \\ \alpha_h^{(C)} \end{bmatrix},$$

and

$$\bar{X}_t^{(C)} = \begin{bmatrix} 0 & \bar{\Theta}_t^{(2-)} & \bar{\Theta}_t^{(3-)} & 0 & \bar{\Theta}_t^{(4-)} & 0 & 0 & 0 \\ 0 & \bar{\Theta}_t^{(1+)} & 0 & \bar{\Theta}_t^{(2+)} & 0 & \bar{\Theta}_t^{(3+)} & 0 & 0 \\ \bar{\Xi}_t^{(2+)} & 0 & 0 & \bar{\Xi}_t^{(3+)} & 0 & 0 & \bar{\Xi}_t^{(4+)} & 0 \\ \bar{\Xi}_t^{(1-)} & 0 & \bar{\Xi}_t^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-)} \\ 0 & \bar{\Theta}_{t-1}^{(2-)} & \bar{\Theta}_{t-1}^{(3-)} & 0 & \bar{\Theta}_{t-1}^{(4-)} & 0 & 0 & 0 \\ 0 & \bar{\Theta}_{t-1}^{(1+)} & 0 & \bar{\Theta}_{t-1}^{(2+)} & 0 & \bar{\Theta}_{t-1}^{(3+)} & 0 & 0 \\ \bar{\Xi}_{t-1}^{(2+)} & 0 & 0 & \bar{\Xi}_{t-1}^{(3+)} & 0 & 0 & \bar{\Xi}_{t-1}^{(4+)} & 0 \\ \bar{\Xi}_{t-1}^{(1-)} & 0 & \bar{\Xi}_{t-1}^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_{t-1}^{(3-)} \end{bmatrix},$$

we write the sample analog as follows:

$$\bar{Y}_t^{(C)} = \bar{X}_t^{(C)} \theta^{(C)}.$$

The C estimator and its estimated asymptotic variance-covariance matrix are defined, replacing symbols $A - r$ and $A[-r]$ by C in the superscripts appended to the variables and statistics regarding the A estimator.

C On uniqueness of estimators

We illustrate the conditions under which the A , B , and C estimators have unique solutions. A slightly older version of Maxima (2025) is used in calculating the determinants appeared in this appendix.

First, we illustrate the conditions for the following two A estimators: those without using $E[\kappa_{it}] = 0$ and without using $E[(1 - y_{i,t-3})' \kappa_{it}] = 0$.

For the former, after swapping the rows and columns of $\bar{X}_t^{(A-3)}$, we obtain the following

7 × 7 block matrix:

$$\begin{bmatrix} 0 & \bar{\Theta}_t^{(2-)} & \bar{\Theta}_t^{(3-)} & 0 & \bar{\Theta}_t^{(4-)} & 0 & 0 \\ \bar{\Xi}_t^{(1-)} & 0 & \bar{\Xi}_t^{(2-)} & 0 & 0 & \bar{\Xi}_t^{(3-)} & 0 \\ \bar{\Theta}_t^{(4+)} & \bar{\Theta}_t^{(2+)} & 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(3+)} \\ 0 & 0 & 0 & \bar{\Xi}_t^{(1++)} & 0 & \bar{\Xi}_t^{(3++)} & \bar{\Xi}_t^{(2++)} \\ \hline 0 & \bar{\Theta}_t^{(2-+)} & \bar{\Theta}_t^{(3-+)} & 0 & \bar{\Theta}_t^{(4-+)} & 0 & 0 \\ \bar{\Xi}_t^{(1-+)} & 0 & \bar{\Xi}_t^{(2-+)} & 0 & 0 & \bar{\Xi}_t^{(3-+)} & 0 \\ \bar{\Theta}_t^{(4++)} & \bar{\Theta}_t^{(2++)} & 0 & 0 & 0 & 0 & \bar{\Theta}_t^{(3++)} \end{bmatrix}.$$

Accordingly, the estimator $\hat{\theta}^{(A[-3])}$ is unique, at least if $\bar{\Lambda}_t^{(A)} \neq 0$ and $\bar{\Xi}_t^{1++} \bar{\Delta}_t^{(A)} \neq 0$, where $\bar{\Lambda}_t^{(A)} = \bar{\Theta}_t^{(3++)} \bar{\Theta}_t^{(4-+)} \bar{\Xi}_t^{(3-+)}$ and

$$\begin{aligned} \bar{\Delta}_t^{(A)} &= \left(\bar{\Theta}_t^{(2-)} - \bar{\Theta}_t^{(2-+)} \frac{\bar{\Theta}_t^{(4-)}}{\bar{\Theta}_t^{(4-+)}} \right) \left(\bar{\Theta}_t^{(4+)} - \bar{\Theta}_t^{(4++)} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_t^{(3++)}} \right) \\ &\times \left(\bar{\Xi}_t^{(2-)} - \bar{\Xi}_t^{(2-+)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_t^{(3-+)}} \right) \\ &+ \left(\bar{\Theta}_t^{(2+)} - \bar{\Theta}_t^{(2++)} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_t^{(3++)}} \right) \left(\bar{\Theta}_t^{(3-)} - \bar{\Theta}_t^{(3-+)} \frac{\bar{\Theta}_t^{(4-)}}{\bar{\Theta}_t^{(4-+)}} \right) \\ &\times \left(\bar{\Xi}_t^{(1-)} - \bar{\Xi}_t^{(1-+)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_t^{(3-+)}} \right). \end{aligned}$$

For the latter, after swapping the rows of $\bar{X}_t^{(A-(3+7))}$, we obtain the following 6 × 6 block matrix:

$$\begin{bmatrix} 0 & \bar{\Theta}_t^{(2-)} & \bar{\Theta}_t^{(3-)} & \bar{\Theta}_t^{(4-)} & 0 & 0 \\ \bar{\Xi}_t^{(1-)} & 0 & \bar{\Xi}_t^{(2-)} & 0 & \bar{\Xi}_t^{(3-)} & 0 \\ \bar{\Theta}_t^{(4+)} & \bar{\Theta}_t^{(2+)} & 0 & 0 & 0 & \bar{\Theta}_t^{(3+)} \\ \hline 0 & \bar{\Theta}_t^{(2-+)} & \bar{\Theta}_t^{(3-+)} & \bar{\Theta}_t^{(4-+)} & 0 & 0 \\ \bar{\Xi}_t^{(1-+)} & 0 & \bar{\Xi}_t^{(2-+)} & 0 & \bar{\Xi}_t^{(3-+)} & 0 \\ \bar{\Theta}_t^{(4++)} & \bar{\Theta}_t^{(2++)} & 0 & 0 & 0 & \bar{\Theta}_t^{(3++)} \end{bmatrix}.$$

Accordingly, the estimator $\hat{\theta}_{[-e]}^{(A[-(3+7)])}$ is unique, at least if $\bar{\Lambda}_t^{(A)} \neq 0$ and $\bar{\Delta}_t^{(A)} \neq 0$.

Second, we illustrate the conditions for the following two B estimators: those without using $E[\lambda_{it}] = 0$ and without using $E[(1 \ y_{i,t-3})' \lambda_{it}] = 0$.

For the former, after swapping the rows and columns of $\bar{X}_t^{(B-1)}$, we obtain the following

7 × 7 block matrix:

$$\begin{bmatrix} 0 & \bar{\Xi}_t^{(2+)} & \bar{\Xi}_t^{(3+)} & 0 & \bar{\Xi}_t^{(4+)} & 0 & 0 \\ \bar{\Theta}_t^{(1+)} & 0 & \bar{\Theta}_t^{(2+)} & 0 & 0 & \bar{\Theta}_t^{(3+)} & 0 \\ \bar{\Xi}_t^{(4-)} & \bar{\Xi}_t^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-)} \\ 0 & 0 & 0 & \bar{\Theta}_t^{(1-+)} & 0 & \bar{\Theta}_t^{(3-+)} & \bar{\Theta}_t^{(2-+)} \\ \hline 0 & \bar{\Xi}_t^{(2\bar{+}\bar{+})} & \bar{\Xi}_t^{(3\bar{+}\bar{+})} & 0 & \bar{\Xi}_t^{(4\bar{+}\bar{+})} & 0 & 0 \\ \bar{\Theta}_t^{(1\bar{+}\bar{+})} & 0 & \bar{\Theta}_t^{(2\bar{+}\bar{+})} & 0 & 0 & \bar{\Theta}_t^{(3\bar{+}\bar{+})} & 0 \\ \bar{\Xi}_t^{(4-+)} & \bar{\Xi}_t^{(2-+)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-+)} \end{bmatrix}.$$

Accordingly, the estimator $\hat{\theta}^{(B[-1])}$ is unique, at least if $\bar{\Lambda}_t^{(B)} \neq 0$ and $\bar{\Theta}_t^{1-+} \bar{\Delta}_t^{(B)} \neq 0$, where $\bar{\Lambda}_t^{(B)} = \bar{\Theta}_t^{(3\bar{+}\bar{+})} \bar{\Xi}_t^{(3-+)} \bar{\Xi}_t^{(4\bar{+}\bar{+})}$ and

$$\begin{aligned} \bar{\Delta}_t^{(B)} &= \left(\bar{\Theta}_t^{(1+)} - \bar{\Theta}_t^{(1\bar{+}\bar{+})} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_t^{(3\bar{+}\bar{+})}} \right) \left(\bar{\Xi}_t^{(2-)} - \bar{\Xi}_t^{(2-+)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_t^{(3-+)}} \right) \\ &\times \left(\bar{\Xi}_t^{(3+)} - \bar{\Xi}_t^{(3\bar{+}\bar{+})} \frac{\bar{\Xi}_t^{(4+)}}{\bar{\Xi}_t^{(4\bar{+}\bar{+})}} \right) \\ &+ \left(\bar{\Theta}_t^{(2+)} - \bar{\Theta}_t^{(2\bar{+}\bar{+})} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_t^{(3\bar{+}\bar{+})}} \right) \left(\bar{\Xi}_t^{(4-)} - \bar{\Xi}_t^{(4-+)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_t^{(3-+)}} \right) \\ &\times \left(\bar{\Xi}_t^{(2+)} - \bar{\Xi}_t^{(2\bar{+}\bar{+})} \frac{\bar{\Xi}_t^{(4+)}}{\bar{\Xi}_t^{(4\bar{+}\bar{+})}} \right). \end{aligned}$$

For the latter, after swapping the rows of $\bar{X}_t^{(B-(1+5))}$, we obtain the following 6 × 6 block matrix:

$$\begin{bmatrix} 0 & \bar{\Xi}_t^{(2+)} & \bar{\Xi}_t^{(3+)} & \bar{\Xi}_t^{(4+)} & 0 & 0 \\ \bar{\Theta}_t^{(1+)} & 0 & \bar{\Theta}_t^{(2+)} & 0 & \bar{\Theta}_t^{(3+)} & 0 \\ \bar{\Xi}_t^{(4-)} & \bar{\Xi}_t^{(2-)} & 0 & 0 & 0 & \bar{\Xi}_t^{(3-)} \\ \hline 0 & \bar{\Xi}_t^{(2\bar{+}\bar{+})} & \bar{\Xi}_t^{(3\bar{+}\bar{+})} & \bar{\Xi}_t^{(4\bar{+}\bar{+})} & 0 & 0 \\ \bar{\Theta}_t^{(1\bar{+}\bar{+})} & 0 & \bar{\Theta}_t^{(2\bar{+}\bar{+})} & 0 & \bar{\Theta}_t^{(3\bar{+}\bar{+})} & 0 \\ \bar{\Xi}_t^{(4-+)} & \bar{\Xi}_t^{(2-+)} & 0 & 0 & 0 & \bar{\Xi}_t^{(3-+)} \end{bmatrix}.$$

Accordingly, the estimator $\hat{\theta}_{[-e]}^{(B[-(1+5)])}$ is unique, at least if $\bar{\Lambda}_t^{(B)} \neq 0$ and $\bar{\Delta}_t^{(B)} \neq 0$.

Third, we illustrate the conditions for the C estimator. $\bar{X}_t^{(C)}$ is a 8 × 8 block matrix as

follows:

$$\left[\begin{array}{cccc|cccc} 0 & \bar{\Theta}_t^{(2-)} & \bar{\Theta}_t^{(3-)} & 0 & \bar{\Theta}_t^{(4-)} & 0 & 0 & 0 \\ 0 & \bar{\Theta}_t^{(1+)} & 0 & \bar{\Theta}_t^{(2+)} & 0 & \bar{\Theta}_t^{(3+)} & 0 & 0 \\ \bar{\Xi}_t^{(2+)} & 0 & 0 & \bar{\Xi}_t^{(3+)} & 0 & 0 & \bar{\Xi}_t^{(4+)} & 0 \\ \bar{\Xi}_t^{(1-)} & 0 & \bar{\Xi}_t^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_t^{(3-)} \\ \hline 0 & \bar{\Theta}_{t-1}^{(2-)} & \bar{\Theta}_{t-1}^{(3-)} & 0 & \bar{\Theta}_{t-1}^{(4-)} & 0 & 0 & 0 \\ 0 & \bar{\Theta}_{t-1}^{(1+)} & 0 & \bar{\Theta}_{t-1}^{(2+)} & 0 & \bar{\Theta}_{t-1}^{(3+)} & 0 & 0 \\ \bar{\Xi}_{t-1}^{(2+)} & 0 & 0 & \bar{\Xi}_{t-1}^{(3+)} & 0 & 0 & \bar{\Xi}_{t-1}^{(4+)} & 0 \\ \bar{\Xi}_{t-1}^{(1-)} & 0 & \bar{\Xi}_{t-1}^{(2-)} & 0 & 0 & 0 & 0 & \bar{\Xi}_{t-1}^{(3-)} \end{array} \right].$$

Accordingly, the estimator $\hat{\theta}^{(C)}$ is unique, at least if $\bar{\Theta}_{t-1}^{(3+)} \bar{\Theta}_{t-1}^{(4-)} \bar{\Xi}_{t-1}^{(3-)} \bar{\Xi}_{t-1}^{(4+)} \neq 0$ and

$$\begin{aligned} & \left(\bar{\Theta}_t^{(1+)} - \bar{\Theta}_{t-1}^{(1+)} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_{t-1}^{(3+)}} \right) \left(\bar{\Theta}_t^{(3-)} - \bar{\Theta}_{t-1}^{(3-)} \frac{\bar{\Theta}_t^{(4-)}}{\bar{\Theta}_{t-1}^{(4-)}} \right) \\ & \times \left(\bar{\Xi}_t^{(1-)} - \bar{\Xi}_{t-1}^{(1-)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_{t-1}^{(3-)}} \right) \left(\bar{\Xi}_t^{(3+)} - \bar{\Xi}_{t-1}^{(3+)} \frac{\bar{\Xi}_t^{(4+)}}{\bar{\Xi}_{t-1}^{(4+)}} \right) \\ & - \left(\bar{\Theta}_t^{(2+)} - \bar{\Theta}_{t-1}^{(2+)} \frac{\bar{\Theta}_t^{(3+)}}{\bar{\Theta}_{t-1}^{(3+)}} \right) \left(\bar{\Theta}_t^{(2-)} - \bar{\Theta}_{t-1}^{(2-)} \frac{\bar{\Theta}_t^{(4-)}}{\bar{\Theta}_{t-1}^{(4-)}} \right) \\ & \times \left(\bar{\Xi}_t^{(2-)} - \bar{\Xi}_{t-1}^{(2-)} \frac{\bar{\Xi}_t^{(3-)}}{\bar{\Xi}_{t-1}^{(3-)}} \right) \left(\bar{\Xi}_t^{(2+)} - \bar{\Xi}_{t-1}^{(2+)} \frac{\bar{\Xi}_t^{(4+)}}{\bar{\Xi}_{t-1}^{(4+)}} \right) \\ & \neq 0. \end{aligned}$$

D Estimators for original parameters

We illustrate the procedures for estimating the original parameters and their variances using the estimation results on the transformed parameters. A highly accessible and easy-to-understand guide to the delta method is provided by Taboga (2021).

First, we illustrate the procedure using the A estimator. A fragment of transformed parameters is written as follows:

$$\begin{aligned} \alpha_a^{(A)} &= \phi_t = \exp(\Delta T D_t), \\ \alpha_b^{(A)} &= \phi_{t+1}^{-1} = \exp(-\Delta T D_{t+1}), \\ \alpha_d^{(A)} &= \phi_t(1 + \delta) = \alpha_a^{(A)} \exp(\gamma). \end{aligned}$$

Accordingly, the original parameters are estimated using the following transformations:

$$\begin{aligned} \gamma &= \log \alpha_d^{(A)} - \log \alpha_a^{(A)}, \\ \Delta T D_t &= \log \alpha_a^{(A)}, \\ \Delta T D_{t+1} &= -\log \alpha_b^{(A)}. \end{aligned}$$

The asymptotic variances of A estimators of γ , ΔTD_t , and ΔTD_{t+1} are estimated using the delta method as follows:

$$\begin{aligned}\widehat{\text{var}}(\hat{\gamma}) &= [1/\hat{\alpha}_d^{(A)} \quad -1/\hat{\alpha}_a^{(A)}] \widehat{\text{Cov}}(\hat{\alpha}_d^{(A)}, \hat{\alpha}_a^{(A)}) [1/\hat{\alpha}_d^{(A)} \quad -1/\hat{\alpha}_a^{(A)}]', \\ \widehat{\text{var}}(\widehat{\Delta TD}_t) &= (1/\hat{\alpha}_a^{(A)})^2 \widehat{\text{var}}(\hat{\alpha}_a^{(A)}), \\ \widehat{\text{var}}(\widehat{\Delta TD}_{t+1}) &= (1/\hat{\alpha}_b^{(A)})^2 \widehat{\text{var}}(\hat{\alpha}_b^{(A)}),\end{aligned}$$

where by adding the circumflex over the symbols of parameters, variance (i.e., var), and variance-covariance matrix (i.e., Cov), their estimators are expressed according to the tradition. If we use $\alpha_e^{(A)}$ instead of $\alpha_d^{(A)}$, we use the transformation $\gamma = -\log \alpha_e^{(A)} + \log \alpha_a^{(A)}$, where $\widehat{\text{Var}}(\hat{\gamma})$ is calculated replacing $\hat{\alpha}_d^{(A)}$ by $\hat{\alpha}_e^{(A)}$.

Second, we illustrate the procedure using the B estimator. A fragment of transformed parameters is written as follows:

$$\begin{aligned}\alpha_a^{(B)} &= \phi_t^{-1} = \exp(-\Delta TD_t), \\ \alpha_b^{(B)} &= \phi_{t+1} = \exp(\Delta TD_{t+1}), \\ \alpha_d^{(B)} &= \phi_t^{-1}(1 + \delta) = \alpha_a^{(B)} \exp(\gamma).\end{aligned}$$

Accordingly, the original parameters are estimated using the following transformations:

$$\begin{aligned}\gamma &= \log \alpha_d^{(B)} - \log \alpha_a^{(B)}, \\ \Delta TD_t &= -\log \alpha_a^{(B)}, \\ \Delta TD_{t+1} &= \log \alpha_b^{(B)}.\end{aligned}$$

The asymptotic variances of B estimators of γ , ΔTD_t , and ΔTD_{t+1} are estimated, replacing a symbol A by B in the superscripts appended to the statistics used in calculating those of A estimators.

Finally, we illustrate the procedure using the C estimator. A fragment of transformed parameters is written as follows:

$$\begin{aligned}\alpha_a^{(C)} &= \phi = \exp(\varphi), \\ \alpha_e^{(C)} &= \phi(1 + \delta) = \alpha_a^{(C)} \exp(\gamma).\end{aligned}$$

Accordingly, the original parameters are estimated using the following transformation:

$$\begin{aligned}\gamma &= \log \alpha_e^{(C)} - \log \alpha_a^{(C)}, \\ \varphi &= \log \alpha_a^{(C)}.\end{aligned}$$

The asymptotic variances of C estimators of γ and φ are estimated as follows:

$$\begin{aligned}\widehat{\text{var}}(\hat{\gamma}) &= [1/\hat{\alpha}_e^{(C)} \quad -1/\hat{\alpha}_a^{(C)}] \widehat{\text{Cov}}(\hat{\alpha}_e^{(C)}, \hat{\alpha}_a^{(C)}) [1/\hat{\alpha}_e^{(C)} \quad -1/\hat{\alpha}_a^{(C)}]', \\ \widehat{\text{var}}(\hat{\varphi}) &= (1/\hat{\alpha}_a^{(C)})^2 \widehat{\text{var}}(\hat{\alpha}_a^{(C)}).\end{aligned}$$

E Estimations of ΔTD_{t-1}

We illustrate the estimations of ΔTD_{t-1} for the following two cases: the cases using both $E[\iota_{i,t-1}] = 0$ and the estimators of $\alpha_a^{(A)}$ and $\alpha_d^{(A)}$ and using both $E[\mu_{i,t-1}] = 0$ and those of $\alpha_a^{(B)}$ and $\alpha_d^{(B)}$. Accordingly, the estimations of ΔTD_{t-1} using the estimators $\hat{\theta}_{[-d]}^{(A[-(1+5)])}$ and $\hat{\theta}_{[-d]}^{(B[-(3+7)])}$ are not available for the former and latter cases, respectively.

For the former case, considering that

$$\iota_{i,t-1} = (1 - y_{i,t-3})(\Theta_{i,t-1}^{(1)} + (1/\alpha_a^{(A)}) \Theta_{i,t-1}^{(2)} + \phi_{t-1} \alpha_a^{(A)} \Theta_{i,t-1}^{(3)} + \phi_{t-1} (\alpha_d^{(A)}/\alpha_a^{(A)}) \Theta_{i,t-1}^{(4)}),$$

the following estimator for ϕ_{t-1} is constructed:

$$\hat{\phi}_{t-1} = -\bar{\Theta}_{t-1}^{(12-)\dagger} / \bar{\Theta}_{t-1}^{(34-)\dagger},$$

where $\bar{\Theta}_{t-1}^{(12-)\dagger} = \hat{\alpha}_a^{(A)} \bar{\Theta}_{t-1}^{(1-)} + \bar{\Theta}_{t-1}^{(2-)}$ and $\bar{\Theta}_{t-1}^{(34-)\dagger} = (\hat{\alpha}_a^{(A)})^2 \bar{\Theta}_{t-1}^{(3-)} + \hat{\alpha}_d^{(A)} \bar{\Theta}_{t-1}^{(4-)}$, with $\bar{\Theta}_{t-1}^{(34-)\dagger} \neq 0$ assumed.

We define

$$\bar{Y}_t^{(A-r)\dagger} = \begin{bmatrix} -\bar{\Theta}_{t-1}^{(12-)\dagger} \\ \bar{Y}_t^{(A-r)} \end{bmatrix}, \quad \hat{\theta}^{(A[-r])\dagger} = \begin{bmatrix} \hat{\phi}_{t-1} \\ \hat{\theta}^{(A[-r])} \end{bmatrix},$$

and

$$\bar{X}_t^{(A-r)\dagger} = \begin{bmatrix} \bar{\Theta}_{t-1}^{(34-)\dagger} & \mathbf{0}' \\ \mathbf{0} & \bar{X}_t^{(A-r)} \end{bmatrix},$$

where if r is replaced by $(3+7)$, $\hat{\theta}^{(A[-r])}$ is replaced by $\hat{\theta}_{[-e]}^{(A[-(3+7)])}$. Then, the estimated asymptotic variance-covariance matrix of $\hat{\theta}^{(A[-r])\dagger}$ is

$$\widehat{\text{Var}}(\hat{\theta}^{(A[-r])\dagger}) = (1/N) \left((\bar{X}_t^{(A-r)\dagger})' \hat{W}_t^{(A-r)\dagger} \bar{X}_t^{(A-r)\dagger} \right)^{-1},$$

where

$$\hat{W}_t^{(A-r)\dagger} = \left((1/N) \sum_{i=1}^N \hat{V}_{it}^{(A-r)\dagger} (\hat{V}_{it}^{(A-r)\dagger})' \right)^{-1}.$$

In this case, the following residual vector is used:

$$\hat{V}_{it}^{(A-r)\dagger} = Y_{it}^{(A-r)\dagger} - X_{it}^{(A-r)\dagger} \hat{\theta}^{(A[-r])\dagger},$$

where $Y_{it}^{(A-r)\dagger}$ and $X_{it}^{(A-r)\dagger}$ are the i th summands in the calculation of the averages $\bar{Y}_t^{(A-r)\dagger}$ and $\bar{X}_t^{(A-r)\dagger}$, respectively.

Given the variance-covariance matrix above, the estimated corrected asymptotic variance of $\hat{\phi}_{t-1}$ is calculated with a generalization of the delta method (see Ruud, 2000, pp. 506–507):

$$\widehat{\text{var}}_c(\hat{\phi}_{t-1}) = \widehat{\text{var}}(\hat{\phi}_{t-1})$$

$$\begin{aligned}
& + 2\bar{J}^{(A)\dagger} [\widehat{\text{cov}}(\hat{\phi}_{t-1}, \hat{\alpha}_a^{(A)}) \quad \widehat{\text{cov}}(\hat{\phi}_{t-1}, \hat{\alpha}_d^{(A)})]' \\
& + \bar{J}^{(A)\dagger} \widehat{\text{Cov}}(\hat{\alpha}_a^{(A)}, \hat{\alpha}_d^{(A)}) (\bar{J}^{(A)\dagger})',
\end{aligned}$$

where $\bar{J}^{(A)\dagger} = [-\bar{\Theta}_{t-1}^{(1-)} + 2\hat{\phi}_{t-1} \hat{\alpha}_a^{(A)} \bar{\Theta}_{t-1}^{(3-)} / \bar{\Theta}_{t-1}^{(34-)\dagger} \quad -\hat{\phi}_{t-1} \bar{\Theta}_{t-1}^{(4-)} / \bar{\Theta}_{t-1}^{(34-)\dagger}]$ and the symbol cov denotes covariance. This estimation for $\hat{\phi}_{t-1}$ is a kind of two-step estimations. The variance correction for two-step estimations dates back to Newey (1984), Pagan (1984), and Murphy and Topel (1985).

Since $\hat{\phi}_{t-1} = \exp(\Delta T D_{t-1})$, the following estimator for $\Delta T D_{t-1}$ is constructed using the estimator for $\hat{\phi}_{t-1}$ above:

$$\widehat{\Delta T D}_{t-1} = \log \hat{\phi}_{t-1}.$$

Then, the corrected asymptotic variance of $\widehat{\Delta T D}_{t-1}$ is estimated as follows:

$$\widehat{\text{var}}_c(\widehat{\Delta T D}_{t-1}) = (1/\hat{\phi}_{t-1})^2 \widehat{\text{var}}_c(\hat{\phi}_{t-1}).$$

For the latter case, considering that

$$\mu_{i,t-1} = y_{i,t-3}(\bar{\Xi}_{i,t-1}^{(1)} + (1/\alpha_a^{(B)}) \bar{\Xi}_{i,t-1}^{(2)} + \phi_{t-1}^{-1} \alpha_a^{(B)} \bar{\Xi}_{i,t-1}^{(3)} + \phi_{t-1}^{-1} (\alpha_d^{(B)}/\alpha_a^{(B)}) \bar{\Xi}_{i,t-1}^{(4)}),$$

the following estimator for ϕ_{t-1}^{-1} is constructed:

$$\hat{\phi}_{t-1}^{-1} = -\bar{\Xi}_{t-1}^{(12+)\dagger} / \bar{\Xi}_{t-1}^{(34+)\dagger},$$

where $\bar{\Xi}_{t-1}^{(12+)\dagger} = \hat{\alpha}_a^{(B)} \bar{\Xi}_{t-1}^{(1+)} + \bar{\Xi}_{t-1}^{(2+)}$ and $\bar{\Xi}_{t-1}^{(34+)\dagger} = (\hat{\alpha}_a^{(B)})^2 \bar{\Xi}_{t-1}^{(3+)} + \hat{\alpha}_d^{(B)} \bar{\Xi}_{t-1}^{(4+)}$, with $\bar{\Xi}_{t-1}^{(34+)\dagger} \neq 0$ assumed.

We define

$$\bar{Y}_t^{(B-r)\dagger} = \begin{bmatrix} -\bar{\Xi}_{t-1}^{(12+)\dagger} \\ \bar{Y}_t^{(B-r)} \end{bmatrix}, \quad \hat{\theta}_t^{(B[-r])\dagger} = \begin{bmatrix} \hat{\phi}_{t-1}^{-1} \\ \hat{\theta}^{(B[-r])} \end{bmatrix},$$

and

$$\bar{X}_t^{(B-r)\dagger} = \begin{bmatrix} \bar{\Xi}_{t-1}^{(34+)\dagger} & \mathbf{0}' \\ \mathbf{0} & \bar{X}_t^{(B-r)} \end{bmatrix},$$

where if r is replaced by $(1+5)$, $\hat{\theta}^{(B[-r])}$ is replaced by $\hat{\theta}_{[-e]}^{(B[-(1+5)])}$. Then, the estimated asymptotic variance-covariance matrix of $\hat{\theta}^{(B[-r])\dagger}$ is

$$\widehat{\text{Var}}(\hat{\theta}^{(B[-r])\dagger}) = (1/N) \left((\bar{X}_t^{(B-r)\dagger})' \hat{W}_t^{(B-r)\dagger} \bar{X}_t^{(B-r)\dagger} \right)^{-1},$$

where

$$\hat{W}_t^{(B-r)\dagger} = \left((1/N) \sum_{i=1}^N \hat{V}_t^{(B-r)\dagger} (\hat{V}_t^{(B-r)\dagger})' \right)^{-1}.$$

In this case, the following residual vector is used:

$$\hat{V}_t^{(B-r)\dagger} = Y_t^{(B-r)\dagger} - X_t^{(B-r)\dagger} \hat{\theta}_t^{(B[-r])\dagger},$$

where $Y_t^{(B-r)\dagger}$ and $X_t^{(B-r)\dagger}$ are the i th summands in the calculation of the averages $\bar{Y}_t^{(B-r)\dagger}$ and $\bar{X}_t^{(B-r)\dagger}$, respectively.

Given the variance-covariance matrix above, the estimated corrected asymptotic variance of $\hat{\phi}_{t-1}^{-1}$ is calculated in the same manner as done in the former case:

$$\begin{aligned} \widehat{\text{var}}_c(\hat{\phi}_{t-1}^{-1}) &= \widehat{\text{var}}(\hat{\phi}_{t-1}^{-1}) \\ &+ 2\bar{J}^{(B)\dagger} [\widehat{\text{cov}}(\hat{\phi}_{t-1}^{-1}, \hat{\alpha}_a^{(B)}) \quad \widehat{\text{cov}}(\hat{\phi}_{t-1}^{-1}, \hat{\alpha}_d^{(B)})]' \\ &+ \bar{J}^{(B)\dagger} \widehat{\text{Cov}}(\hat{\alpha}_a^{(B)}, \hat{\alpha}_d^{(B)}) (\bar{J}^{(B)\dagger})', \end{aligned}$$

where $\bar{J}^{(B)\dagger} = [-(\bar{\Xi}_{t-1}^{(1+)} + 2\hat{\phi}_{t-1}^{-1} \hat{\alpha}_a^{(B)} \bar{\Xi}_{t-1}^{(3+)}) / \bar{\Xi}_{t-1}^{(34+)\dagger} \quad - \hat{\phi}_{t-1}^{-1} \bar{\Xi}_{t-1}^{(4+)} / \bar{\Xi}_{t-1}^{(34+)\dagger}]$.

Since $\hat{\phi}_{t-1}^{-1} = \exp(-\Delta T D_{t-1})$, the following estimator for $\Delta T D_{t-1}$ is constructed using the estimator for $\hat{\phi}_{t-1}^{-1}$ above:

$$\widehat{\Delta T D}_{t-1} = -\log \hat{\phi}_{t-1}^{-1}.$$

Then, the corrected asymptotic variance of $\widehat{\Delta T D}_{t-1}$ is estimated as follows:

$$\widehat{\text{var}}_c(\widehat{\Delta T D}_{t-1}) = (1/\hat{\phi}_{t-1}^{-1})^2 \widehat{\text{var}}_c(\hat{\phi}_{t-1}^{-1}).$$

F Monte Carlo results for transformed parameters

Tables F.1, F.2, and F.3 present Monte Carlo results for the transformed parameters of A –(3+7), B –(1+5), and C estimators, respectively. A calculation indicates that convergence rates of the estimators are root- N . Further, the histograms and Q-Q plots depict asymptotic normality of the estimators.

Table F.1: Monte Carlo results of A-(3+7) estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
α_a	1.22140	1.22054	0.03373	0.03401	-0.00086	0.03374
α_b	1.34986	1.35700	0.17472	0.17675	0.00715	0.17486
α_c	0.90484	0.89574	0.18863	0.19061	-0.00910	0.18884
α_d	3.32012	3.31682	0.13089	0.13229	-0.00329	0.13093
α_f	0.49659	0.50078	0.07233	0.07297	0.00420	0.07245
α_g	0.33287	0.32916	0.08321	0.08397	-0.00371	0.08329
ϕ_{t-1}	1.49182	1.49210	0.03049	0.03026	0.00028	0.03049
$N = 50$ million						
α_a	1.22140	1.22201	0.01512	0.01509	0.00060	0.01513
α_b	1.34986	1.34816	0.07906	0.07833	-0.00169	0.07908
α_c	0.90484	0.90672	0.08485	0.08426	0.00188	0.08487
α_d	3.32012	3.32132	0.05938	0.05888	0.00120	0.05940
α_f	0.49659	0.49621	0.03239	0.03219	-0.00037	0.03240
α_g	0.33287	0.33328	0.03750	0.03720	0.00040	0.03751
ϕ_{t-1}	1.49182	1.49212	0.01326	0.01345	0.00029	0.01327
$N = 100$ million						
α_a	1.22140	1.22133	0.01060	0.01066	-0.00008	0.01060
α_b	1.34986	1.35105	0.05529	0.05535	0.00119	0.05530
α_c	0.90484	0.90356	0.05933	0.05954	-0.00128	0.05934
α_d	3.32012	3.31928	0.04183	0.04160	-0.00083	0.04184
α_f	0.49659	0.49723	0.02272	0.02276	0.00064	0.02273
α_g	0.33287	0.33217	0.02634	0.02630	-0.00070	0.02635
ϕ_{t-1}	1.49182	1.49205	0.00971	0.00951	0.00022	0.00972

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation. 2) In each replication, ϕ_{t-1} (where $t = 7$) is estimated with the two-step estimation and its standard error is calculated with the corrected variance. 3) The superscript of parameters is omitted.

Table F.2: Monte Carlo results of $B-(1+5)$ estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
α_a	0.81873	0.82018	0.02338	0.02355	0.00145	0.02342
α_b	0.74082	0.72954	0.13768	0.13891	-0.01128	0.13815
α_c	1.10517	1.11926	0.17366	0.17512	0.01409	0.17423
α_d	2.22554	2.23291	0.09948	0.10050	0.00737	0.09975
α_f	0.27253	0.26708	0.06190	0.06216	-0.00546	0.06214
α_g	0.40657	0.41238	0.06929	0.06987	0.00581	0.06953
ϕ_{t-1}^{-1}	0.67032	0.66975	0.01367	0.01358	-0.00057	0.01368
$N = 50$ million						
α_a	0.81873	0.81848	0.01034	0.01038	-0.00025	0.01035
α_b	0.74082	0.74120	0.06114	0.06088	0.00039	0.06114
α_c	1.10517	1.10451	0.07700	0.07679	-0.00066	0.07700
α_d	2.22554	2.22593	0.04443	0.04395	0.00039	0.04443
α_f	0.27253	0.27230	0.02726	0.02718	-0.00024	0.02726
α_g	0.40657	0.40645	0.03073	0.03060	-0.00012	0.03073
ϕ_{t-1}^{-1}	0.67032	0.67003	0.00601	0.00605	-0.00029	0.00602
$N = 100$ million						
α_a	0.81873	0.81886	0.00732	0.00734	0.00013	0.00732
α_b	0.74082	0.73943	0.04296	0.04306	-0.00139	0.04298
α_c	1.10517	1.10684	0.05424	0.05432	0.00167	0.05427
α_d	2.22554	2.22663	0.03068	0.03107	0.00109	0.03070
α_f	0.27253	0.27177	0.01928	0.01923	-0.00077	0.01929
α_g	0.40657	0.40731	0.02160	0.02164	0.00074	0.02161
ϕ_{t-1}^{-1}	0.67032	0.67021	0.00429	0.00427	-0.00011	0.00429

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation. 2) In each replication, ϕ_{t-1}^{-1} (where $t = 7$) is estimated with the two-step estimation and its standard error is calculated with the corrected variance. 3) The superscript of parameters is omitted.

Table F.3: Monte Carlo results of C estimator

parameter	true	mean	sd	se	bias	rmse
$N = 10$ million						
α_a	1.34986	1.35740	0.17626	0.17754	0.00754	0.17642
α_b	0.74082	0.74020	0.01038	0.01033	-0.00062	0.01040
α_c	1.82212	1.82125	0.32887	0.33150	-0.00087	0.32887
α_d	0.54881	0.54384	0.09981	0.10020	-0.00497	0.09994
α_e	3.66930	3.67009	0.08917	0.08978	0.00079	0.08918
α_f	0.49659	0.49855	0.13884	0.14172	0.00196	0.13885
α_g	2.01375	2.01784	0.22723	0.23292	0.00409	0.22726
α_h	0.27253	0.27731	0.03375	0.03355	0.00478	0.03408
$N = 50$ million						
α_a	1.34986	1.35146	0.07817	0.07861	0.00160	0.07819
α_b	0.74082	0.74084	0.00460	0.00460	0.00003	0.00460
α_c	1.82212	1.82256	0.14842	0.14787	0.00044	0.14842
α_d	0.54881	0.54797	0.04416	0.04436	-0.00084	0.04417
α_e	3.66930	3.66915	0.04027	0.04006	-0.00015	0.04027
α_f	0.49659	0.49809	0.06198	0.06274	0.00151	0.06200
α_g	2.01375	2.01140	0.10187	0.10274	-0.00236	0.10189
α_h	0.27253	0.27340	0.01406	0.01431	0.00087	0.01409
$N = 100$ million						
α_a	1.34986	1.35104	0.05740	0.05555	0.00118	0.05741
α_b	0.74082	0.74087	0.00326	0.00325	0.00005	0.00326
α_c	1.82212	1.82255	0.10714	0.10460	0.00043	0.10714
α_d	0.54881	0.54825	0.03240	0.03134	-0.00056	0.03241
α_e	3.66930	3.66907	0.02906	0.02833	-0.00023	0.02906
α_f	0.49659	0.49780	0.04551	0.04434	0.00121	0.04553
α_g	2.01375	2.01133	0.07440	0.07260	-0.00242	0.07444
α_h	0.27253	0.27315	0.01047	0.01007	0.00062	0.01049

Notes: 1) In each replication, the last five cross-sections y_{i4}, \dots, y_{i8} for $i = 1, \dots, N$ are used for the estimation. 2) The superscript of parameters is omitted.

G Linear restrictions of transformed parameters

We illustrate the linear restrictions tested using the Wald test, after obtaining the estimation results from using the A -(3+4), B -(1+5), and C estimators.

The three linear restrictions among the logarithms of the transformed parameters for the A -(3+4) and B -(1+5) estimators are

$$\begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 2 & -1 & 0 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \log \alpha_a \\ \log \alpha_b \\ \log \alpha_c \\ \log \alpha_d \\ \log \alpha_f \\ \log \alpha_g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

while the six ones for the C estimator are

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ -2 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \log \alpha_a \\ \log \alpha_b \\ \log \alpha_c \\ \log \alpha_d \\ \log \alpha_e \\ \log \alpha_f \\ \log \alpha_g \\ \log \alpha_h \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

where the superscripts of parameters are suppressed.

In addition, if we estimate the model with time trends using the A -(3+4) and B -(1+5) estimators, we have the following four linear restrictions for the Wald test:

$$\begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 \\ 3 & 0 & 0 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} \log \alpha_a \\ \log \alpha_b \\ \log \alpha_c \\ \log \alpha_d \\ \log \alpha_f \\ \log \alpha_g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$