1-2017

Determining individual or time effects in panel data models

Xun LU

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Determining Individual or Time Effects in Panel Data Models*

Xun Lu[a] and Liangjun Su[b]

[a]Department of Economics, Hong Kong University of Science & Technology

[b]School of Economics, Singapore Management University, Singapore

January 6, 2017

Abstract

In this paper we propose a jackknife method to determine individual and time effects in linear panel data models. We first show that when both the serial and cross-sectional correlation among the idiosyncratic error terms are weak, our jackknife method can pick up the correct model with probability approaching one (w.p.a.1). In the presence of moderate or strong degree of serial correlation, we modify our jackknife criterion function and show that the modified jackknife method can also select the correct model w.p.a.1. We conduct Monte Carlo simulations to show that our new methods perform remarkably well in finite samples. We apply our methods to study (*i*) the crime rates in North Carolina, (*ii*) the determinants of saving rates across countries, and (*iii*) the relationship between guns and crime rates in the U.S.

Key words: Consistency, Cross-validation, Dynamic panel, Information Criterion, Jackknife, Individual effect, Time effect.

JEL Classification: C23, C33, C51, C52.

*Su gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under Grant MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. All errors are the authors' sole responsibilities. Address correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Phone: +65 6828 0386; e-mail: ljsu@smu.edu.sg.

1 Introduction

Individual effects and time effects are often used in panel data models to model unobserved individual or time heterogeneity (see, e.g., Arellano (2003), Baltagi (2013), Hsiao (2014), and Wooldridge (2010) for a review on panel data models). The goal of this paper is to provide practical methods to determine whether to include individual effects, or time effects, or both in linear panel data models. Specifically, we consider the following four models:

$$\begin{aligned}
&\text{Model 1:} && y_{it} = \beta' x_{it} + u_{it}, \\
&\text{Model 2:} && y_{it} = \beta' x_{it} + \alpha_i + u_{it}, \\
&\text{Model 3:} && y_{it} = \beta' x_{it} + \lambda_t + u_{it}, \\
&\text{Model 4:} && y_{it} = \beta' x_{it} + \alpha_i + \lambda_t + u_{it},
\end{aligned}$$

where $i = 1, ..., N$, $t = 1, ..., T$, x_{it} is a $k \times 1$ vector of regressors that may include lagged dependent variables, α_i is an individual effect, λ_t is a time effect, and u_{it} is an idiosyncratic error term. We will treat α_i's and λ_t's as fixed parameters to be estimated. For clarity, we assume that x_{it} contains the constant term in all models and impose restrictions on α_i or/and λ_t in Models 2-4 to achieve identification for the fixed effects. Specifically, we assume that

$$\sum_{i=1}^{N} \alpha_i = 0 \text{ in Model 2,} \tag{1.1}$$

$$\sum_{t=1}^{T} \lambda_t = 0 \text{ in Model 3, and} \tag{1.2}$$

$$\sum_{i=1}^{N} \alpha_i = 0 \text{ and } \sum_{t=1}^{T} \lambda_t = 0 \text{ in Model 4.} \tag{1.3}$$

The above identification restrictions greatly facilitate the asymptotic analysis in this paper and make it straightforward to extend the methodology developed here to multi-dimensional panel data models.[1]

We propose a jackknife or leave-one-out cross-validation (CV) method to select the correct model.[2] There are several advantages of our jackknife method in the context of determining fixed effects. First, the new method is general and easy to implement. It does not require the choice of any tuning parameter that is implicitly used in all information-criterion-based methods (e.g., a Bayesian information criterion (BIC) specifies the penalty term to be proportional to $\ln(NT)/(NT)$, which works as a tuning parameter). Second, we assume that the cross-section

[1] For our method discussed below, different identification restrictions, e.g., assuming $\alpha_N = 0$ in Model 2 and $\lambda_T = 0$ in Model 3, produce identical results.

[2] Throughout the paper, we use Jackknife and CV interchangeably. Jackknife is widely used in model selection and model averaging (see, e.g., Allen (1974), Stone (1974), Geisser (1974), Wahba and Wold (1975), Li (1987), Andrews (1991), Hansen and Racine (2012), and Lu and Su (2015)).

dimension (N) and time dimension (T) pass to infinity simultaneously. But the relative rate between N and T can be arbitrary. For example, T can be much slower than N such as $T \asymp \ln(N)$. This implies that our method can be applied to the typical case in micro-econometrics where T is much smaller than N. Third, our CV method can be applied to both static and dynamic panel models. We show that when serial correlation and cross-sectional dependence in the error term are absent or weak, our CV method can choose the correct model with probability approaching one (w.p.a.1).[3] Fourth, we propose a modified CV method that is robust to strong serial correlation in the static panel models. We show that the modified CV can select the correct model w.p.a.1. in the presence of strong serial correlation. Fifth, our jackknife method can be easily extended to nonlinear panels and to multi-level panels where the determination of different fixed effects is also imperative.

In the literature, there exist several tests for testing for the presence of fixed effects in two dimensional panel data models. Most of the tests focus on short static panel models. Let σ_α^2 and σ_λ^2 be the variances of α_i and λ_t, respectively. Under the normality assumption, Breusch and Pagan (1980, BP hereafter) propose a Lagrange multiplier (LM) test for testing the null hypothesis: $H_{01}: \sigma_\alpha^2 = 0$ and $\sigma_\lambda^2 = 0$. The BP test can also be applied to test the null hypotheses that $H_{02}: \sigma_\alpha^2 = 0$ (assuming $\sigma_\lambda^2 = 0$) and that $H_{03}: \sigma_\lambda^2 = 0$ (assuming $\sigma_\alpha^2 = 0$) (see, e.g., Baltagi, 2013 for a discussion). Honda (1985) shows that BP test is actually robust to the non-normality and also modifies the test to a one-sided test. Baltagi, Chang, and Li (1992, BCL hereafter) modify the one-side test based on the results of Gourieroux, Holly, and Monfort (1982). BCL also propose conditional LM tests for testing $H_{04}: \sigma_\alpha^2 = 0$ (allowing $\sigma_\lambda^2 > 0$) and $H_{05}: \sigma_\lambda^2 = 0$ (allowing $\sigma_\alpha^2 > 0$). Moulton and Randolph (1989) consider the ANOVA F-test. All the tests discussed above assume that the error terms $\{u_{it}, t = 1, ..., T\}$ are not serially correlated. Bera, Sosa-Escudero, and Yoon (2001) propose an LM test that allows serial correlation in the error term. Recently, Wu and Li (2014) propose Hausman-type tests for testing H_{01}, H_{04} and H_{05} by comparing the variances of the error terms at different robust levels. Wu and Zhu (2012) extend the Hausman-type tests to short dynamic panel models.

Potentially, these tests can be used to determine the correct model. For example, we can test H_{01}, H_{04}, and H_{05} sequentially. However, there are several limitations of the approach based on the hypothesis testing. First, to determine the correct model, three separate tests need to be implemented sequentially. This involves the multiple testing issue and it is unclear how to choose an appropriate nominal level.[4] In addition, in finite samples, it could occur that H_{01} is rejected, while neither H_{04} nor H_{05} is rejected, in which case it is difficult to decide the correct model.

[3] We only allow serial correlation in static panel models. For dynamic panel data models (e.g., panel AR(1) model), the serial correlation in the error terms (e.g., AR(1) errors) will cause the error terms to be correlated with the lagged dependent variables. We do not address the endogeneity issue in this paper.

[4] There is a large literature on the multiple testing issue for controlling the family-wise error rate (FWER). See, e.g., Romano, Shaikh and Wolf (2010) for a review. However, to the best of our knowledge, there is no discussion on how to address this issue in the context of determining fixed effects.

Second, the existing tests are designed for short panels (i.e., T is fixed), and it is unclear how the tests behave when T also goes to infinity. We consider large panels where N and T go to infinity simultaneously and we allow the relative rates of N and T to be arbitrary. Third, except Wu and Zhu (2012), most existing tests do not apply to dynamic panel models, i.e., the regressors cannot contain any lagged dependent variables.

Alternatively, we can consider certain information criteria (IC) such as AIC and BIC. However, to the best of our knowledge, there is no theoretical analysis of AIC or BIC in the context of determining fixed effects in panel data. When all four models are allowed, a careful analysis indicates that AIC is always inconsistent and BIC is consistent in the special case where N and T pass to infinity at the same rate. In Monte Carlo simulations we compare our jackknife method with AIC and BIC, and find that our jackknife method generally outperforms this IC-based approach.

The rest of the paper is structured as follows. In Section 2, we propose the jackknife and the modified jackknife method and study their asymptotic properties. Section 3 reports Monte Carlo simulation results and compares our new methods with IC-based methods for both static and dynamic panel data generating processes. In Section 4, we provide three empirical applications. In the first application, we study the crime rates in North Carolina and find that Model 4 is the correct model. The second application is about the determinants of saving rates across countries and our methods select Model 2. In the third application, we investigate the relationship between guns and crime rates in the U.S. and we determine that Model 4 is the correct model.

Notation. For an $m \times n$ real matrix A, we denote its transpose as A' and its Frobenius norm as $\|A\|$ ($\equiv [\text{tr}(AA')]^{1/2}$) where $\equiv$ means "is defined as". Let $P_A \equiv A\left(A'A\right)^{-1}A'$ and $M_A \equiv I_m - P_A$, where I_m denotes an $m \times m$ identity matrix. When $A = \{a_{ij}\}$ is symmetric, we use $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ to denote its maximum and minimum eigenvalues, respectively. The operator $\xrightarrow{P}$ denotes convergence in probability. We use $(N,T) \to \infty$ to denote that N and T pass to infinity simultaneously.

2 Methodology and Asymptotic Theory

In this section, we first introduce the jackknife method to determine individual or time effects in panel data models and then study the consistency of our jackknife estimator. To allow for strong degree of serial correlation we also propose a modified jackknife criterion function and justify its asymptotic validity.

2.1 Methodology

Let $x_i = (x_{i1}, ..., x_{iT})'$ and $X = (x_1', ..., x_N')'$. Define y_i, u_i, Y, and U analogously. To facilitate the presentation, we define the following dummy matrices:

$$D_\alpha = \begin{pmatrix} I_{N-1} \\ -\iota_{N-1}' \end{pmatrix} \otimes \iota_T, \ D_\lambda = \iota_N \otimes \begin{pmatrix} I_{T-1} \\ -\iota_{T-1}' \end{pmatrix}, \text{ and } D_{\alpha\lambda} = (D_\alpha, D_\lambda),$$

where ι_a is an $a \times 1$ vector of ones for any integer $a \geq 1$. To unify the notation, we write

$$X^{(1)} = X, \ X^{(2)} = (X, D_\alpha), \ X^{(3)} = (X, D_\lambda), \text{ and } X^{(4)} = (X, D_\alpha, D_\lambda).$$

We use $x_{it}^{(m)\prime}$ to denote a typical row of $X^{(m)}$ such that $X^{(m)} = (x_{11}^{(m)}, ..., x_{1T}^{(m)}, ..., x_{N1}^{(m)}, ..., x_{NT}^{(m)})$ for $m = 1, 2, 3, 4$. Similarly, we use $d_{\alpha,it}'$, $d_{\lambda,it}'$, and $d_{\alpha\lambda,it}'$ to denote a typical row of D_α, D_λ, and $D_{\alpha\lambda}$, respectively. Then we can rewrite Models 1-4 as follows:

$$\begin{aligned}
&\text{Model 1:} \quad && y_{it} = \beta' x_{it} + u_{it} \equiv \beta^{(1)\prime} x_{it}^{(1)} + u_{it}, \\
&\text{Model 2:} \quad && y_{it} = \beta' x_{it} + \underline{\alpha}' d_{\alpha,it} + u_{it} \equiv \beta^{(2)\prime} x_{it}^{(2)} + u_{it}, \\
&\text{Model 3:} \quad && y_{it} = \beta' x_{it} + \underline{\lambda}' d_{\lambda,it} + u_{it} \equiv \beta^{(3)\prime} x_{it}^{(3)} + u_{it}, \\
&\text{Model 4:} \quad && y_{it} = \beta' x_{it} + \underline{\alpha}' d_{\alpha,it} + \underline{\lambda}' d_{\lambda,it} + u_{it} \equiv \beta^{(4)\prime} x_{it}^{(4)} + u_{it},
\end{aligned}$$

where $\underline{\alpha} = (\alpha_1, ..., \alpha_{N-1})'$, $\underline{\lambda} = (\lambda_1, ..., \lambda_{T-1})'$, $\beta^{(1)} = \beta$, $\beta^{(2)} = (\beta', \underline{\alpha}')'$, $\beta^{(3)} = (\beta', \underline{\lambda}')'$, and $\beta^{(4)} = (\beta', \underline{\alpha}', \underline{\lambda}')'$. Note that we have imposed the identification conditions in (1.1)-(1.3) for Models 2-4 in the above representation. In matrix notation, we can write these models simply as

$$\begin{aligned}
&\text{Model 1:} \quad && Y = X\beta + U = X^{(1)}\beta^{(1)} + U, \\
&\text{Model 2:} \quad && Y = X\beta + D_\alpha \underline{\alpha} + U = X^{(2)}\beta^{(2)} + U, \\
&\text{Model 3:} \quad && Y = X\beta + D_\lambda \underline{\lambda} + U = X^{(3)}\beta^{(3)} + U, \\
&\text{Model 4:} \quad && Y = X\beta + D_\alpha \underline{\alpha} + D_\lambda \underline{\lambda} + U = X^{(4)}\beta^{(4)} + U.
\end{aligned}$$

Note that Model 1 is nested in Models 2-4, both Models 2 and 3 are nested in Model 4, and $D_\alpha' D_\lambda = 0$. These observations greatly simplify the asymptotic analysis in this paper.

The OLS estimator of $\beta^{(m)}$ based on all observations $\{(y_{it}, x_{it}^{(m)}) : 1 \leq i \leq N, \ 1 \leq t \leq T\}$ is given by

$$\hat{\beta}^{(m)} = \left(X^{(m)\prime} X^{(m)}\right)^{-1} X^{(m)\prime} Y \text{ for } m = 1, 2, 3, 4. \tag{2.1}$$

We also consider the leave-one-out estimator of $\beta^{(m)}$ with the (i, t)th observation deleted:

$$\hat{\beta}_{it}^{(m)} = \left(X^{m\prime} X^{(m)} - x_{it}^{(m)} x_{it}^{(m)\prime}\right)^{-1} \left(X^{(m)\prime} Y - x_{it}^{(m)} y_{it}\right) \text{ for } m = 1, 2, 3, 4, \tag{2.2}$$

where $i = 1, ..., N$, $t = 1, ..., T$. Define the out-of-sample predicted value of y_{it} as $\hat{y}_{it}^{(m)} = \hat{\beta}_{it}^{(m)\prime} x_{it}^{(m)}$. Our jackknife method is based on the following leave-one-out CV function

$$CV(m) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(y_{it} - \hat{y}_{it}^{(m)}\right)^2 \text{ for } m = 1, 2, 3, 4. \tag{2.3}$$

Let

$$\hat{m} = \underset{1 \leq m \leq 4}{\operatorname{argmin}}\; CV(m). \tag{2.4}$$

Under some regularity conditions, we will show that w.p.a.1, $\hat{m}$ is given by m when Model m is the true model.

2.2 Asymptotic theory under weak serial and cross-sectional correlations

Let $\bar{u}_{i\cdot} = T^{-1}\sum_{t=1}^{T} u_{it}$, $\bar{u}_{\cdot t} = N^{-1}\sum_{i=1}^{N} u_{it}$, and $\bar{u}_{\cdot\cdot} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}$. Let $\bar{x}_{i\cdot}$, $\bar{x}_{\cdot t}$, and $\bar{x}_{\cdot\cdot}$ be defined analogously. Define

$$\hat{Q} = \frac{1}{NT}X'X \text{ and } \hat{Q}_{D_\xi} = \frac{1}{NT}X'M_{D_\xi}X \text{ for } D_\xi = D_\alpha,\ D_\lambda, \text{and } D_{\alpha\lambda}.$$

Let C denote a generic large positive constant whose value may vary across lines.

To proceed, we make the following set of assumptions.

Assumption A.1. (i) $E(u_{it}) = 0$, $\max_{1\leq i\leq N, 1\leq t\leq T} E\left(u_{it}^2\right) \leq C$, and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 \xrightarrow{P} \bar{\sigma}_u^2 > 0$.

(ii) $\max_{1\leq i\leq N, 1\leq t\leq T} \|x_{it}\| = o_P(C_{NT})$ where $C_{NT} = \min(N^{3/4}T^{1/4}, N^{1/4}T^{3/4})$.

(iii) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(\|x_{it}\|^2 u_{it}^2\right) = O(1)$.

(iv) $\bar{u}_{\cdot\cdot} = O_P\left((NT)^{-1/2}\right)$, $\bar{x}_{\cdot\cdot} = O_P(1)$, and $\frac{1}{NT}X'U = O_P\left((NT)^{-1/2}\right)$.

(v) There exist positive constants $\underline{c}_Q$ and $\bar{c}_Q$ such that

$$P\left(\underline{c}_Q \leq \lambda_{\min}\left(\hat{Q}_{D_\xi}\right) \leq \lambda_{\max}\left(\hat{Q}\right) \leq \bar{c}_Q\right) \to 1$$

for $D_\xi = D_\alpha$, D_λ, and $D_{\alpha\lambda}$.

(vi) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\alpha_i = o_P(1)$ and $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}\lambda_t = o_P(1)$ when Model 2, 3, or 4 is true and applicable.

Assumption A.2. (i) $\frac{T}{N}\sum_{i=1}^{N}(\bar{u}_{i\cdot})^2 \xrightarrow{P} \bar{\sigma}_{u1}^2 > 0$.

(ii) $\frac{N}{T}\sum_{t=1}^{T}(\bar{u}_{\cdot t})^2 \xrightarrow{P} \bar{\sigma}_{u2}^2 > 0$.

(iii) $\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot} = O_P\left(T^{-1} + (NT)^{-1/2}\right)$.

(iv) $\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{\cdot t}\bar{u}_{\cdot t} = O_P\left(N^{-1} + (NT)^{-1/2}\right)$.

Assumption A.3. (i) If Model 2 is the true model, there exist positive constants $c_{\alpha,X}$ and c_{α,X_λ} such that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\alpha_i - x_{it}'\left(X'X\right)^{-1}X'D_\alpha\underline{\alpha}\right]^2 \xrightarrow{P} c_{\alpha,X} \quad > \quad 0, \text{ and} \tag{2.5}$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\alpha_i - x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_\alpha\underline{\alpha}\right]^2 \xrightarrow{P} c_{\alpha,X_\lambda} \quad > \quad 0. \tag{2.6}$$

(ii) If Model 3 is the true model, there exist positive constants $c_{\lambda,X}$ and c_{λ,X_α} such that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_t - x_{it}'\left(X'X\right)^{-1}X'D_\lambda\underline{\lambda}\right]^2 \xrightarrow{P} c_{\lambda,X} \quad > \quad 0, \text{ and} \tag{2.7}$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_t - x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}D_\lambda\underline{\lambda}\right]^2 \xrightarrow{P} c_{\lambda,X_\alpha} \quad > \quad 0. \tag{2.8}$$

(iii) If Model 4 is the true model, there exist positive constants $c_{\alpha\lambda,X}$, c_{α,X_λ}, and c_{λ,X_α} such that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\alpha_i + \lambda_t - x_{it}'\left(X'X\right)^{-1}X'\left(D_\alpha\underline{\alpha} + D_\lambda\underline{\lambda}\right)\right]^2 \xrightarrow{P} c_{\alpha\lambda,X} > 0 \tag{2.9}$$

and both (2.6) and (2.8) hold.

Assumptions A.1(i)-(iii) impose weak conditions on $\{u_{it}\}$ and $\{x_{it}\}$, which can be verified under various primitive conditions (see, e.g., Baltagi (2013), Hsiao (2014), and Wooldridge (2010)). For example, if $E\left\|x_{it}\right\|^4$ is uniformly bounded, then by the Markov inequality and dominated convergence theorem (DCT) we can readily show that $\max_{1\leq i\leq N,1\leq t\leq T}\left\|x_{it}\right\| = o_P((NT)^{1/4})$, which is sufficient for A.1(ii). Similarly, a sufficient condition for Assumption A.1(iii) to hold is that both $E\left\|x_{it}\right\|^4$ and $E\left(u_{it}^4\right)$ are uniformly bounded. Assumption A.1(iv) is also weak and commonly assumed in panel data models in the absence of endogeneity. In particular, we permit x_{it} to contain lagged dependent variables so that dynamic panel data models are allowed. Assumption A.1(v) specifies the usual identification conditions for the OLS or fixed effects (FE) estimation of Models 1-4. For example, the condition that $\lambda_{\min}(\hat{Q}_{D_\alpha})$ is bounded below from 0 requires that x_{it} should not contain any time-invariant regressor beyond a constant term; it is allowed to contain a constant term because we have imposed the identification constraint that $\sum_{i=1}^N \alpha_i = 0$. Similarly, the condition that $\lambda_{\min}(\hat{Q}_{D_\lambda})$ is bounded below from 0 requires that x_{it} should not contain any individual-invariant regressor beyond a constant term; it is allowed to contain a constant term because we have imposed the identification constraint that $\sum_{t=1}^T \lambda_t = 0$. On the surface, this condition rules out the inclusion of any time-invariant regressor in Model 2, individual-invariant regressor in Model 3, and both types of regressors in Model 4. If x_{it} contains such regressors, they should be removed from Models 2-4 and then we can redefine $x_{it}^{(m)}$ for $m = 2, 3, 4$ with such regressors removed. Then the asymptotic analysis below will continue to hold. Assumption A.1(vi) essentially imposes conditions on the interactions between the idiosyncratic error terms and the individual and time effects, whenever applicable, in Models 2-4. A sufficient condition for it to hold is that both $\{u_{it}\alpha_i\}$ and $\{u_{it}\lambda_t\}$ have zero mean and follow some version of weak law of large numbers. The zero mean condition is commonly assumed in the panel data literature. Note that we allow the individual effects α_i and time effects λ_t to be random in the true model (if present) even if we treat them as fixed parameters in the estimation procedure.

Assumption A.2(i) requires that $\{u_{it}, t \geq 1\}$ be weakly serially dependent such that $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \sum_{s=1}^T E\left(u_{it}u_{is}\right)$ has a finite limit. For example, the latter condition is satisfied by the Davydov inequality if $\{u_{it}, t \geq 1\}$ is strong mixing with finite $(2+\delta)$ −th moment and mixing coefficients $\alpha_i(\cdot)$ such that $\alpha_i(\tau) = \tau^{-\gamma_i}$ for some $\gamma_i > (2+\delta)/\delta$; see, e.g., Bosq (1998, pp.19-20) or the online supplement of Su, Shi, and Phillips (2016). Similarly, Assumption A.2(ii) requires that $\{u_{it}, i \geq 1\}$ be weakly cross-sectionally dependent such that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} E(u_{it}u_{jt})$ has a finite limit. Assumption A.2(iii)-(iv) can be verified under both weak serial and cross-sectional correlations by the Chebyshev inequality and it is easily met in the absence of both serial and cross-sectional correlations. If there is no serial correlation among $\{u_{it}, t \geq 1\}$, then $\bar{\sigma}_{u1}^2 = \bar{\sigma}_u^2$; if there is no cross-sectional correlation among $\{u_{it}, i \geq 1\}$, then $\bar{\sigma}_{u2}^2 = \bar{\sigma}_u^2$. When serial correlation is present, $\bar{\sigma}_{u1}^2$ is generally different from $\bar{\sigma}_u^2$; when cross-sectional correlation is present, $\bar{\sigma}_{u2}^2$ is generally different from $\bar{\sigma}_u^2$.

Assumption A.3 specifies conditions to ensure that the underfitted models will never be chosen asymptotically. The interpretations of the conditions in (2.5)-(2.9) are easy. For example, when Model 2 is the true model, both Models 1 and 3 are underfitted. In this case, (2.5) and (2.6) require that the individual effects α_i, when stacked into an $NT \times 1$ vector, should not lie in the column space spanned by the regressor matrix X in Model 1 and $X^{(3)}$ in Model 3, respectively. Similarly, when Model 4 is the true model, Models 1, 2, and 3 are all underfitted. In this case, (2.9) requires that $\alpha_i + \lambda_t$, when stacked into an $NT \times 1$ vector, should not lie in the column space spanned by the regressor matrix X in Model 1, (2.8) requires that the time effects λ_t should not lie in the column space spanned by $X^{(2)}$ in Model 2, and (2.6) requires that the individual effects α_i should not lie in the column space spanned by $X^{(3)}$ in Model 3.

It is worth mentioning that we allow for both cross-sectional and serial dependence of unknown form in $\{(x_{it}, u_{it})\}$ despite the fact that some of the results derived below need further constraints. We do not need identical distributions or homoskedasticity along either the cross-section dimension or the time dimension, neither do we need to assume mean or covariance stationarity along either dimension. In this sense, we say our results below are applicable to a variety of linear panel data models in practice.

Given Assumptions A.1-A.3, we are ready to state our first main result.

Theorem 2.1 *Suppose that Assumptions A.1-A.3 hold. Suppose that* $\max\left(\bar{\sigma}_{u1}^2, \bar{\sigma}_{u2}^2\right) < 2\bar{\sigma}_u^2$, *where* $\bar{\sigma}_{u1}^2, \bar{\sigma}_{u2}^2$, *and* $\bar{\sigma}_u^2$ *are defined in Assumptions 2(i), 2(ii), and 1(i), respectively. Then*

$$P(\hat{m} = m \mid \textit{Model } m \textit{ is the true model}) \to 1 \textit{ as } (N,T) \to \infty \textit{ for } m = 1, ..., 4.$$

Remark 1. The proof of Theorem 2.1 is given in the supplementary appendix. To appreciate the above result, we outline the main idea that underlines our proof. When Model 1 is true, all the other models are overfitted, and we can show that $P(CV(1) < CV(m)) \to 1$ for $m = 2, 3, 4$

by showing that

$$\begin{aligned}
T\left[CV\left(2\right)-CV\left(1\right)\right] \xrightarrow{P} 2\bar{\sigma}_u^2-\bar{\sigma}_{u1}^2 &> 0,\\
N\left[CV\left(3\right)-CV\left(1\right)\right] \xrightarrow{P} 2\bar{\sigma}_u^2-\bar{\sigma}_{u2}^2 &> 0,\\
\left(N\wedge T\right)\left[CV\left(4\right)-CV\left(1\right)\right] \xrightarrow{P} 2\left(1+c\right)\bar{\sigma}_u^2-\left(\bar{\sigma}_{u1}^2+c\bar{\sigma}_{u2}^2\right)1\left\{c_1\geq 1\right\}-\left(c\bar{\sigma}_{u1}^2+\bar{\sigma}_{u2}^2\right)1\left\{c_1<1\right\} &> 0,
\end{aligned}$$

where $c=\lim_{(N,T)\to\infty}\left(\frac{N}{T}\wedge\frac{T}{N}\right)$, and $c_1=\lim_{(N,T)\to\infty}\frac{N}{T}$, and $a\wedge b=\min\left(a,b\right)$. When Model 2 is true, Models 1 and 3 are underfitted, Model 4 is overfitted, and we can show that $P\left(CV\left(2\right)<CV\left(m\right)\right)\to 1$ for $m=1,3,4$ by showing that

$$\begin{aligned}
CV\left(1\right)-CV\left(2\right) \xrightarrow{P} c_{\alpha,X} &> 0,\\
CV\left(3\right)-CV\left(2\right) \xrightarrow{P} c_{\alpha,X_\lambda} &> 0,\\
N\left[CV\left(4\right)-CV\left(2\right)\right] \xrightarrow{P} 2\bar{\sigma}_u^2-\bar{\sigma}_{u2}^2 &> 0.
\end{aligned}$$

When Model 3 is true, Models 1 and 2 are underfitted, Model 4 is overfitted, and we can show that $P\left(CV\left(3\right)<CV\left(m\right)\right)\to 1$ for $m=1,2,4$ by showing that

$$\begin{aligned}
CV\left(1\right)-CV\left(3\right) \xrightarrow{P} c_{\lambda,X} &> 0,\\
CV\left(2\right)-CV\left(3\right) \xrightarrow{P} c_{\lambda,X_\alpha} &> 0,\\
T\left[CV\left(4\right)-CV\left(3\right)\right] \xrightarrow{P} 2\bar{\sigma}_u^2-\bar{\sigma}_{u1}^2 &> 0.
\end{aligned}$$

When Model 4 is true, all other models are underfitted, and we can show that $P\left(CV\left(4\right)<CV\left(m\right)\right)\to 1$ for $m=1,2,3$ by showing that

$$\begin{aligned}
CV\left(1\right)-CV\left(4\right) \xrightarrow{P} c_{\alpha\lambda,X} &> 0,\\
CV\left(2\right)-CV\left(4\right) \xrightarrow{P} c_{\lambda,X_\alpha} &> 0,\\
CV\left(3\right)-CV\left(4\right) \xrightarrow{P} c_{\alpha,X_\lambda} &> 0.
\end{aligned}$$

As a result, $CV\left(m\right)$ has the minimal value among $\left\{CV\left(l\right),l=1,...,4\right\}$ asymptotically only when Model m is the true model.

Remark 2. Theorem 2.1 indicates that we can choose the correct model w.p.a.1 as $(N,T)\to\infty$. In other words, our jackknife method can choose the correct model consistently as long as the serial or cross-sectional correlation among the error terms is not strong enough to overtake the average noise level as represented by $\bar{\sigma}_u^2$. As remarked above, the additional condition $\max\left(\bar{\sigma}_{u1}^2,\bar{\sigma}_{u2}^2\right)<2\bar{\sigma}_u^2$ would be automatically satisfied in the absence of both serial and cross-sectional correlation among the idiosyncratic error terms. Note that the above result does not have any restriction on the degree of serial or cross-sectional correlation among $\{x_{it}\}$ as long as Assumptions A.1(ii)-(v) are satisfied. More importantly, we do not need any relative rate condition on how N and T pass to infinity. In fact, our theory allows $T=O\left(\ln N\right)$ such that our method may be applied to micro panels when T is typically small in comparison with N.

Remark 3. To see when the above additional condition can be met in Theorem 2.1, we focus on the case where $\{u_{it}, t \geq 1\}$ follows a covariance-stationary AR(1) process with mean zero and variance σ_u^2 for each i. Let $\rho \in (-1, 1)$ denote the AR(1) coefficient. Then by straightforward calculations,

$$\begin{aligned} \frac{T}{N}\sum_{i=1}^{N} E\left(\bar{u}_{i\cdot}\right)^2 &= \frac{1}{T}\sum_{t=1}^{T} E\left(u_{it}^2\right) + \frac{2}{T}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} E\left(u_{it}u_{is}\right) \\ &= \sigma_u^2 + \frac{2\sigma_u^2}{T}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \rho^{s-t} \\ &= \sigma_u^2\left(1 + \frac{2\sigma_u^2}{T}\sum_{t=1}^{T-1}\frac{\rho\left(1-\rho^{T-t+1}\right)}{1-\rho}\right) \\ &\to \sigma_u^2\left(1 + \frac{2\rho}{1-\rho}\right) = \bar{\sigma}_{u1}^2. \end{aligned}$$

In this case, $\bar{\sigma}_u^2 = \sigma_u^2$ and $\bar{\sigma}_{u1}^2 < 2\bar{\sigma}_u^2$ provided $\rho < \frac{1}{3}$. Similarly, if $\{u_{it}, i \geq 1\}$ has mean zero and variance σ_u^2 for each i, t such that $\text{Corr}(u_{it}, u_{jt}) = \rho^{|i-j|}$ for all i, j, t for some $\rho \in (-1, 1)$, then

$$\frac{N}{T}\sum_{i=1}^{T} E\left(\bar{u}_{\cdot t}\right)^2 \to \sigma_u^2\left(1 + \frac{2\rho}{1-\rho}\right) = \bar{\sigma}_{u2}^2$$

and $\bar{\sigma}_{u2}^2 < 2\bar{\sigma}_u^2$ provided $\rho < \frac{1}{3}$.

The above calculations indicate that the serial or cross-sectional correlation among the error terms cannot be moderately large in order for our jackknife method to work. In the next subsection, we consider the relaxation of such conditions. Since there is typically no natural ordering among the individual units, we focus on the relaxation on the serial dependence along the time dimension and propose a modified jackknife criterion function to handle strong or moderately large degree of serial correlation.

2.3 A modified jackknife criterion function

In this subsection, we consider the panel data model with serially correlated errors and propose a modified version of the jackknife criterion function. We assume that the error process $\{u_{it}, t \geq 1\}$ can be approximated by an AR(p) process:

$$u_{it} = \rho_1 u_{i,t-1} + \rho_2 u_{i,t-2} + ... + \rho_p u_{i,t-p} + v_{it} = \boldsymbol{\rho}' \underline{u}_{i,t-1} + v_{it}, \tag{2.10}$$

where $i = 1, ..., N,\ t = p+1, ..., T,\ \boldsymbol{\rho} = \left(\rho_1, ..., \rho_p\right)'$ is a vector of unknown parameters, $\underline{u}_{i,t-1} = (u_{i,t-1}, ..., u_{i,t-p})'$, and v_{it} is an innovation term.

Let $\hat{u}_{it}^{(m)} = y_{it} - \hat{\beta}^{(m)\prime} x_{it}^{(m)}$ for $m = 1, 2, 3, 4$. We propose to estimate the AR(p) coefficients based on the residuals from Model 4 (the largest model), i.e., we run the following regression

$$\hat{u}_{it}^{(4)} = \rho_1 \hat{u}_{i,t-1}^{(4)} + \rho_2 \hat{u}_{i,t-2}^{(4)} + ... + \rho_p \hat{u}_{i,t-p}^{(4)} + \tilde{v}_{it} = \boldsymbol{\rho}' \underline{\hat{u}}_{i,t-1}^{(4)} + \tilde{v}_{it}, \tag{2.11}$$

where $i = 1, ..., N$, $t = p+1, ..., T$, $\underline{\hat{u}}_{i,t-1}^{(4)} = (\hat{u}_{i,t-1}^{(4)}, ..., \hat{u}_{i,t-p}^{(4)})'$, and $\tilde{v}_{it} = (\hat{u}_{it}^{(4)} - u_{it}) + \boldsymbol{\rho}'(\underline{u}_{i,t-1} - \underline{\hat{u}}_{i,t-1}^{(4)}) + v_{it}$. Let $\hat{\boldsymbol{\rho}} = (\hat{\rho}_1, \hat{\rho}_2, ...\hat{\rho}_p)'$ denote the OLS estimator of $\boldsymbol{\rho}$ in the above regression. Let $\underline{y}_{i,t-1} = (y_{i,t-1}, ..., y_{i,t-p})'$ and $\underline{\hat{y}}_{i,t-1}^{(m)} = (\hat{y}_{i,t-1}^{(m)}, ..., \hat{y}_{i,t-p}^{(m)})'$. We modify the CV criterion function as

$$CV^*(m) = \frac{1}{N(T-p)} \sum_{t=p+1}^{T} \sum_{i=1}^{N} \left[\left(y_{it} - \hat{\boldsymbol{\rho}}' \underline{y}_{i,t-1} \right) - \left(\hat{y}_{it}^{(m)} - \hat{\boldsymbol{\rho}}' \underline{\hat{y}}_{i,t-1}^{(m)} \right) \right]^2. \tag{2.12}$$

Let

$$\tilde{m} = \underset{1 \leq m \leq 4}{\operatorname{argmin}} \, CV^*(m). \tag{2.13}$$

Ideally, when Model m is correctly specified, $(y_{it} - \hat{\boldsymbol{\rho}}' \underline{y}_{i,t-1}) - (\hat{y}_{it}^{(m)} - \hat{\boldsymbol{\rho}}' \underline{\hat{y}}_{i,t-1}^{(m)})$ will approximate the true innovation term v_{it}. As long as there is no serial correlation among $\{v_{it}\}$ or the serial correlation is weak, $\tilde{m}$ is given by m w.p.a.1. when Model m is the true model.

Let

$$\Phi(L) = 1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p,$$

where L is the lag operator. Similarly, $\Phi(1) = 1 - \rho_1 - \rho_2 - \cdots \rho_p$. Let $\breve{x}_{it}^{(m)} = \Phi(L) x_{it}^{(m)}$ for $t = p+1, ..., T$ and $m = 1, 2, 3, 4$. Note that $\breve{x}_{it}^{(1)} = \Phi(L) x_{it} \equiv \breve{x}_{it}$. Let $\bar{v}_{i\cdot} = T_p^{-1} \sum_{t=p+1}^{T} v_{it}$ for $i = 1, ...N$, and $\bar{v}_{\cdot t} = N^{-1} \sum_{i=1}^{N} v_{it}$ for $t = p+1, ..., N$, where $T_p = T - p$.

To state the next result, we add the following set of assumptions.

Assumption A.4. (i) All the roots of $\Phi(z)$ lie outside the unit circle.

(ii) $E(v_{it}) = 0$, $\max_{1 \leq i \leq N, p+1 \leq t \leq T} E(v_{it}^2) \leq C$, and $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it}^2 \xrightarrow{P} \bar{\sigma}_v^2 > 0$.

(iii) $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=p+1}^{T} E\left(\|x_{it}\|^2 v_{it}^2 \right) = O(1)$.

(iv) $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \zeta_{it} v_{it} = O_P\left((NT)^{-1/2} \right)$ for $\zeta_{it} = 1$, x_{it}, $x_{i,t-j}$, and $u_{i,t-j}$ where $j = 1, ..., p$.

(v) $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it} \alpha_i = o_P(1)$ and $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it} \lambda_{t-j} = o_P(1)$ for $j = 0, 1, ..., p$ when Model 2, 3, or 4 is true and applicable.

Assumption A.5. (i) $\frac{T_p}{N} \sum_{i=1}^{N} (\bar{v}_{i\cdot})^2 \xrightarrow{P} \bar{\sigma}_{v1}^2 > 0$.

(ii) $\frac{N}{T_p} \sum_{t=p+1}^{T} (\bar{v}_{\cdot t})^2 \xrightarrow{P} \bar{\sigma}_{v2}^2 > 0$.

(iii) $\frac{1}{N} \sum_{i=1}^{N} \bar{x}_{i\cdot} \bar{v}_{i\cdot} = O_P\left(T^{-1} + (NT)^{-1/2} \right)$.

(iv) $\frac{1}{T_p} \sum_{t=p+1}^{T} \bar{x}_{\cdot t} \bar{v}_{\cdot t} = O_P\left(N^{-1} + (NT)^{-1/2} \right)$.

Assumption A.6. (i) If Model 2 is the true model, there exist positive constants $c_{\alpha,X}^*$ and c_{α,X_λ}^* such that

$$\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left[\Phi(1) \alpha_i - \breve{x}_{it}' \left(X'X \right)^{-1} X' D_\alpha \underline{\alpha} \right]^2 \xrightarrow{P} c_{\alpha,X}^* \quad > \quad 0, \text{ and} \tag{2.14}$$

$$\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \left[\Phi(1) \alpha_i - \breve{x}_{it}^{(3)\prime} \left(X^{(3)\prime} X^{(3)} \right)^{-1} X^{(3)\prime} D_\alpha \underline{\alpha} \right]^2 \xrightarrow{P} c_{\alpha,X_\lambda}^* \quad > \quad 0. \tag{2.15}$$

(ii) If Model 3 is the true model, there exist positive constants $c^*_{\lambda,X}$ and c^*_{λ,X_α} such that

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\Phi\left(L\right)\lambda_t-\breve{x}_{it}'\left(X'X\right)^{-1}X'D_\lambda\underline{\lambda}\right]^2\overset{P}{\to}c^*_{\lambda,X}\quad>\quad 0,\text{ and} \tag{2.16}$$

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\Phi\left(L\right)\lambda_t-\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}D_\lambda\underline{\lambda}\right]^2\overset{P}{\to}c^*_{\lambda,X_\alpha}\quad>\quad 0. \tag{2.17}$$

(iii) If Model 4 is the true model, there exist positive constants $c^*_{\alpha\lambda,X}$, c^*_{α,X_λ}, and c^*_{λ,X_α} such that

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\left(\Phi\left(1\right)\alpha_i+\Phi\left(L\right)\lambda_t\right)-\breve{x}_{it}'\left(X'X\right)^{-1}X'\left(D_\alpha\underline{\alpha}+D_\lambda\underline{\lambda}\right)\right]^2\overset{P}{\to}c^*_{\alpha\lambda,X}>0 \tag{2.18}$$

and both (2.15) and (2.17) hold.

Assumption A.4(i) rules out unit root or explosive processes for $\{u_{it},t\geq 1\}$. Assumption A.4(ii)-(v) parallels Assumption A.1(i), (iii)-(iv) and (v). Assumption A.5(i)-(iv) parallels Assumption A.2(i)-(iv). Assumption A.6(i)-(iii) is analogous to Assumption A.3(i)-(iii).

Theorem 2.2 *Suppose that Assumptions A.1-A.2 and A.4-A.6 hold. Suppose that* $\max(\bar{\sigma}^2_{v1},\bar{\sigma}^2_{v2})<2\bar{\sigma}^2_v$. *Then*

$$P\left(\tilde{m}=m\mid \textit{Model } m \textit{ is the true model}\right)\to 1 \textit{ as } (N,T)\to\infty \textit{ for } m=1,...,4.$$

Remark 4. Theorem 2.2 indicates that the modified jackknife criterion function helps us to select the correct model w.p.a.1 as $(N,T)\to\infty$ under the weak side condition $\max(\bar{\sigma}^2_{v1},\bar{\sigma}^2_{v2})<2\bar{\sigma}^2_v$. Where there is no serial correlation among $\{u_{it},t\geq 1\}$ such that $\Phi\left(1\right)=\Phi\left(L\right)=1$ and $u_{it}=v_{it}$, then $\bar{\sigma}^2_{v1}=\bar{\sigma}^2_{u1}=\bar{\sigma}^2_u=\bar{\sigma}^2_v$ and $\bar{\sigma}^2_{v2}=\bar{\sigma}^2_{u2}$. This implies that the result in Theorem 2.2 coincides with that in Theorem 2.1 in this case. If there is no serial or cross-sectional correlation among $\{v_{it}\}$, then $\bar{\sigma}^2_{v1}=\bar{\sigma}^2_{v2}=\bar{\sigma}^2_v$ and $\max(\bar{\sigma}^2_{v1},\bar{\sigma}^2_{v2})<2\bar{\sigma}^2_v$ is automatically satisfied.

Remark 5. In the above analysis, we run the pooled AR(p) regression for $\hat{u}^{(4)}_{it}$. A close examination of the proof of Theorem 2.2 indicates that only the consistency of the pooled OLS estimator $\hat{\boldsymbol{\rho}}$ is used. Alternatively, one can allow heterogeneity in both the order of autoregression and its coefficients. In this case, we use p_i and $\boldsymbol{\rho}_i$, $i=1,...,N$, to denote the order and individual coefficients in the autoregressive models and run the AR(p_i) regression for $\{\hat{u}^{(4)}_{it},t\geq 1\}$ to estimate $\boldsymbol{\rho}_i$ by $\hat{\boldsymbol{\rho}}_i$ for $i=1,...,N$. Then we can modify the jackknife criterion function to be

$$CV^*\left(m\right)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T-p_i}\sum_{t=p_i+1}^{T}\left[\left(y_{it}-\hat{\boldsymbol{\rho}}_i'\underline{y}_{i,t-1}\right)-\left(\hat{y}^{(m)}_{(it)}-\hat{\boldsymbol{\rho}}_i'\underline{\hat{y}}^{(m)}_{i,t-1}\right)\right]^2.$$

Accordingly, we can modify Assumptions A.4-A.6 and establish a result similar to that in Theorem 2.2.

Remark 6. Alternatively, we can rewrite the original model by including p lagged y_{it} and p lagged x_{it} (excluding the constant) as additional (pk) regressors via the standard Cochrane–Orcutt procedure. Take Model 4 as an example. Let $\mathring{x}_{it}$ be the x_{it} excluding the constant term, i.e., $x_{it} = (1, \mathring{x}_{it}')'$. Correspondingly, let $\beta = (\beta_1, \mathring{\beta}')'$. Then, Model 4

$$y_{it} = \beta' x_{it} + \alpha_i + \lambda_t + u_{it} = (\beta_1, \mathring{\beta}')(1, \mathring{x}_{it}')' + \alpha_i + \lambda_t + u_{it}$$

can be rewritten as

$$\begin{aligned} y_{it} &= \left(1 - \rho_1 - ... - \rho_p\right) \beta_1 + \mathring{\beta}' \mathring{x}_{it} + \rho_1 y_{i,t-1} + ... + \rho_p y_{i,t-p} - \left(\rho_1 \mathring{\beta}' \mathring{x}_{i,t-1} + ... + \rho_p \mathring{\beta}' \mathring{x}_{i,t-p}\right) \\ &\quad + \left(1 - \rho_1 - ... - \rho_p\right) \alpha_i + \left(\lambda_t - \rho_1 \lambda_{t-1} - ... - \rho_p \lambda_{t-p}\right) + v_{it} \\ &= \tilde{\beta}' \tilde{x}_{it} + \tilde{\alpha}_i + \tilde{\lambda}_t + v_{it}, \end{aligned}$$

where $\tilde{x}_{it} = \left(1, \mathring{x}_{it}', \ y_{i,t-1}, \ ... \ , \ y_{i,t-p}, \ \mathring{x}_{i,t-1}', \ ..., \ \mathring{x}_{it-p}'\right)'$, $\tilde{\beta}$ is the new vector of regression coefficients, $\tilde{\alpha}_i = \left(1 - \rho_1 - ... - \rho_p\right) \alpha_i$ and $\tilde{\lambda}_t = \left(\lambda_t - \rho_1 \lambda_{t-1} - ... - \rho_p \lambda_{t-p}\right)$. With the new regressor $\tilde{x}_{it}$ replacing x_{it}, we can continue to apply the jackknife criterion function $CV(m)$ as in Section 2.1.

Remark 7. Here we impose an AR(p) structure on the error term. In practice, $\{u_{it}, t \geq 1\}$ do not need to follow the AR(p) process exactly. Note that our original jackknife method in Section 2.1 works in the presence of weak serial correlation. Hence, here it is sufficient to reduce and control the serial correlation among $\{u_{it}, t \geq 1\}$.

3 Monte Carlo Simulations

In this section, we conduct Monte Carlo simulations to examine the finite sample performance of our jackknife method and compare it with various information criteria (IC). We consider the following three different cases: (*i*) static panel models with possibly serially correlated errors, (*ii*) dynamic panel models without exogenous regressors and (*iii*) dynamic panel models with exogenous regressors.

3.1 Implementation

As a comparison, we consider the commonly used information criterion (IC): AIC and BIC, though to the best of our knowledge, there is no theoretical analysis of AIC and BIC in the context of determining fixed effects. Here the number of parameters involved depends on N and T and goes to infinity, thus the standard theory of AIC and BIC is not directly applicable here.

For Model m, $m = 1, 2, 3, 4$, define the in-sample residual as $\hat{u}_{it}^{(m)} = y_{it} - \hat{\beta}^{(m)\prime} x_{it}^{(m)}$. Then AIC and BIC for Model m are defined respectively as

$$\begin{aligned} AIC(m) &= \ln\left(\left(\hat{\sigma}^{(m)}\right)^2\right) + \frac{2k^{(m)}}{NT}, \\ BIC(m) &= \ln\left(\left(\hat{\sigma}^{(m)}\right)^2\right) + \frac{\log(NT)\, k^{(m)}}{NT}, \end{aligned}$$

where $\left(\hat{\sigma}^{(m)}\right)^2 = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\hat{u}_{it}^{(m)}\right)^2$ and $k^{(m)}$ is the dimension of $x_{it}^{(m)}$ in the mth model. Specifically, $k^{(1)} = k$, $k^{(2)} = k + N - 1$, $k^{(3)} = k + T - 1$ and $k^{(4)} = k + N + T - 2$. We also consider the modified BIC as

$$BIC_2(m) = \ln\left(\left(\hat{\sigma}^{(m)}\right)^2\right) + \frac{\log(\log(NT))\, k^{(m)}}{NT}.$$

We choose the model by minimizing the above three ICs.[5]

For static panel models, we consider CV (defined in (2.3)) and CV* (defined in (2.12)). To take into account the possible serial correlation, we also apply CV to the augmented regression with additional p lagged y_{it} and p lagged x_{it} (excluding the constant), as discussed in Remark 6 above. We denote it as CV**. For dynamic panel models, we only consider CV, as serial correlation can cause the endogenous problem and in general is not allowed in dynamic panel models. For all the simulations, we consider different combinations of N and T : $(N, T) = (10, 10)$, $(10, 50)$, $(50, 10)$ and $(50, 50)$. The number of replications is 1000.

3.2 Static panel models

We consider the following static data generating processes (DGPs):

$$\begin{array}{ll} \text{DGP 1.1: } y_{it} = 1 + x_{it} + u_{it} & \text{DGP 1.2: } y_{it} = 1 + x_{it} + \alpha_i + u_{it} \\ \text{DGP 1.3: } y_{it} = 1 + x_{it} + \lambda_t + u_{it} & \text{DGP 1.4: } y_{it} = 1 + x_{it} + \alpha_i + \lambda_t + u_{it} \end{array},$$

where $x_{it} = 1 + \alpha_i + \lambda_t + \xi_{it}$ and α_i, λ_t and ξ_{it} are mutually independent $N(0, 1)$ random variables. The error term u_{it} is generated as

$$u_{it} = \rho u_{i,t-1} + v_{it},$$

where v_{it} is a $N(0, 1)$ random variable, and ρ takes different values: 0, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{3}{4}$. Here the true models corresponding to DGPs 1.1-1.4 are Models 1-4, respectively.

Tables 1A, 1B, 1C, 1D and 1E present the simulation results for $\rho = 0$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{3}{4}$, respectively. When $\rho = 0$, i.e., there is no serial correlation in the error term, our CV performs

[5] Following the standard analysis on the consistency of IC, we can show the following results: (1) BIC and BIC_2 are consistent in selecting the individual or time effects under the restrictive condition that N and T pass to infinity at the same rate; (2) the AIC is never consistent; and (3) neither BIC nor BIC_2 is consistent in general when N and T pass to infinity at different rates.

best. For example, even when $N = 10$ and $T = 10$, our CV can choose the correct model with a probability close to 95%. The performance of AIC is also good and comparable to that of CV. CV* and CV**, which are robust to possible serial correlation, perform slightly worse than CV in the absence of serial correlation. The performance of BIC is poor. For example, when the true model is Model 4 and $(N, T) = (10, 50)$, BIC can only choose the correct model with a probability of 3.5%. BIC_2 outperforms BIC, but still underperforms CV and AIC.

When $\rho = \frac{1}{4}$, i.e., there is weak serial correlation in the error term, our CV* and CV** perform best overall, as suggested by our theory. Between CV* and CV**, it is not apparent which one dominates. For example, when the true model is Model 1, CV** outperforms CV*, but when the true model is Model 2, CV* outperforms CV**. CV also performs reasonably well, as our theory suggests that CV can consistently select the correct model when the serial correlation is weak ($\rho < \frac{1}{3}$ for this DGP). The performance of AIC is slightly worse than that of CV. Both BIC (e.g., when the true model is Model 3 or 4) and BIC_2 (e.g., when the true model is Model 3) perform poorly.

$\rho = \frac{1}{3}$ is an interesting case, as $\rho = \frac{1}{3}$ is the cut-off point for CV to work. In the discussion following Theorem 2.1, we show that when $\rho = \frac{1}{3}$, $\bar{\sigma}_{u1}^2 = 2\bar{\sigma}_u^2$, thus the key condition $\max\left(\bar{\sigma}_{u1}^2, \bar{\sigma}_{u2}^2\right) < 2\bar{\sigma}_u^2$ is violated. In our proof, we show that in this case, when the true model is Model 1, $T\left[CV(2) - CV(1)\right] \xrightarrow{P} 0$ and when the true model is Model 3, $T\left[CV(4) - CV(3)\right] \xrightarrow{P} 0$. This suggests that CV cannot distinguish Model 1 and Model 2 when the true model is Model 1 and cannot distinguish Model 3 and Model 4 when the true model is Model 3. Our simulations confirm the theoretical analysis. For example, when the true model is Model 1 and $(N, T) = (50, 50)$, CV selects Model 1 and Model 2 with probabilities of 55.7% and 44.3%, respectively. In this case, CV, AIC, BIC and BIC_2 all break down. However, both CV* and CV**, which explicitly take serial correlation into account, perform well, as suggested by our theory. For example, when $(N, T) = (50, 50)$, both CV* and CV** can select the correct model with a probability close to 100%. For this DGP, CV* slightly outperforms CV** as a whole.

When the series correlation is high, such as $\rho = \frac{1}{2}$ and $\frac{3}{4}$, the performances of CV, AIC, BIC and BIC_2 are all poor. In general, CV* and CV** perform well, especially when the sample is large. For this DGP, CV* outperforms CV** in general. For example, when $(N, T) = (50, 50)$ and $\rho = \frac{1}{2}$ or $\frac{3}{4}$, CV* can choose the correct model with a probability close to 100%. However, when the true model is Model 4 and $(N, T) = (50, 50)$, $\rho = \frac{3}{4}$, CV** can only choose the correct model with a probability of 49.1% . This seems to suggest that when serial correlation is high, a large sample is required for CV** to work well.

To examine the effect of model misspecification, in Table 4A, we compare the mean squared errors (MSEs) of the estimator of the slope coefficient ($\beta = 1$) using different models when $\rho = 0$.[6] It is clear that for this DGP, the correct model achieves the smallest MSE. For example, when the true model is Model 1 and $(N, T) = (10, 10)$, the MSE based on Model 4 is about 3.5 times

[6] The results for $\rho = \frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{3}{4}$ are avaiable upon request.

as large as that based on Model 1.

In sum, for static panel models, when there is no serial correlation or serial correlation is low, CV, CV*, CV** and AIC all work well. In the absence of serial correlation, CV is the best performer. When serial correlation is high, only CV* and CV** work and CV* generally outperforms CV**.

3.3 Dynamic panel models without exogenous regressors

We consider the following dynamic panel DGPs:

$$\begin{array}{ll} \text{DGP 2.1: } y_{it} = 1 + \beta y_{i,t-1} + u_{it} & \text{DGP 2.2: } y_{it} = 1 + \beta y_{i,t-1} + \alpha_i + u_{it} \\ \text{DGP 2.3: } y_{it} = 1 + \beta y_{i,t-1} + \lambda_t + u_{it} & \text{DGP 2.4: } y_{it} = 1 + \beta y_{i,t-1} + \alpha_i + \lambda_t + u_{it} \end{array},$$

where α_i, λ_t and u_{it} are mutually independent $N(0,1)$ random variables and β takes different values: $\frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$.

Tables 2A, 2B, and 2C report the simulations results for $\beta = \frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$, respectively. For most cases, our CV can select the correct method with a high probability and dominates other methods. Despite its inconsistency, AIC performs slightly worse than CV. For example, when the true model is Model 1, $\beta = \frac{1}{2}$, $(N,T) = (10,10)$, CV and AIC choose the correct model with probabilities of 84.4% and 79.6%, respectively. The performance of BIC is poor in many cases. For example, when the true model is Model 2, $\beta = \frac{1}{2}$, and $(N,T) = (50,10)$, BIC selects the correct model with zero probability. The performance of BIC_2 is better than that of BIC, but still worse than those of CV and AIC in general.

Table 4B shows the MSEs of estimator of β based on the four models when $\beta = \frac{3}{4}$.[7] We consider both the non-bias corrected estimator and bias corrected estimator. For the bias correction, we adopt the half panel jackknife method as proposed in Dhaene and Jochmans (2015). For both types of estimators, the estimator based on the true model has the smallest MSE. For example, when true model is Model 1 and $(N,T) = (10,10)$, the MSEs of the non-bias corrected estimator based on Model 4 is about 10 times as large as that based on Model 1, and the MSE of the bias corrected estimator based on Model 4 is about 5 times as large as that based on Model 1.

[7] The results for $\beta = \frac{1}{4}$ and $\frac{1}{2}$ are avaiable upon request.

3.4 Dynamic panel models with exogenous regressors

We consider the following dynamic panel DGPs with 5 exogenous regressors:

$$\text{DGP 3.1: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2x_{it,j} + u_{it},$$

$$\text{DGP 3.2: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2x_{it,j} + \alpha_i + u_{it},$$

$$\text{DGP 3.3: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2x_{it,j} + \lambda_t + u_{it},$$

$$\text{DGP 3.4: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2x_{it,j} + \alpha_i + \lambda_t + u_{it},$$

where $x_{it,1} = 1+\alpha_i+\lambda_t+ \xi_{it}$, and $x_{it,2}$, $x_{it,3}$, $x_{it,4}$, $x_{it,5}$, α_i, λ_t, u_{it} and ξ_{it} are mutually independent $N(0,1)$ random variables, and β takes different values: $\frac{1}{4}, \frac{1}{2}$ and $\frac{3}{4}$. Here the number of regressors is $k = 7$ (including the constant).

Table 3A, 3B and 3C represent the frequency of the model selected for $\beta = \frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$, respectively. The simulation results are similar to those in the dynamic models without exogenous regressors. In general, our CV performs best, followed by AIC. Both CV and AIC can select the correct model with a high probability, especially when the sample size is large. For example, when $(N, T) = (50, 50)$, the correct probabilities are all close to 100%. BIC performs poorly when the true model is Model 2 or Model 4. BIC_2 outperforms BIC, but still underperforms CV and AIC.

4 Empirical Applications

In this section we consider three empirical applications that illustrate the usefulness of our method in selecting individual or time effects in panel data models.

4.1 Application I: Crime rates in North Carolina

Cornwell and Trumbull (1994) study the crime rates using the panel data on 90 counties in North Carolina over the period 1981 – 1987. The vector of explanatory variables x_{it} includes: (1) the probability of arrest, measured by the ratio of arrests to offences, (2) the probability of conviction given arrest, measured by the ratio of convictions to arrests, (3) the probability of a prison sentence given a conviction, measured by the proportion of total convictions resulting in prison sentences, (4) the average prison sentence in days, (5) the number of police per capita, (6) the population density, measured by the county population divided by the county land area, (7) the percentage of young male, measured by the proportion of the county's population that is male and between the ages of 15 and 24, and $(8-16)$ the average weekly wage in the county in the following nine industries: (i) construction, (ii) transportation, utilities and communication, (iii) wholesale and retail trade, (*iv*) finance, insurance and real estate, (*v*) services, (*vi*) manufacturing, (*vii*) federal government, (*viii*) state government and (*ix*) local government. All the variables are in logarithm. Hence we have a static panel with $N = 90$, $T = 7$ and $k = 17$ (including the constant). The same dataset is also used in Baltagi (2006) and Wu and Li (2014).

Table 5 presents the values of AIC, BIC, BIC_2, CV, CV^*, and CV^{**}. All these methods determine that Model 4 (i.e., including both individual and time fixed effects) is the correct model.

4.2 Application II: Cross-country saving rates

Su, Shi, and Phillips (2016) use a dynamic panel data model to study the determinants of savings rates. Following Edwards (1996), they let y_{it} be the ratio of savings to GDP for country i in year t, and let x_{it} include (*i*) its CPI-based inflation rate, (*ii*) its real interest rate, (*iii*) its per capita GDP growth rate and (*iv*) its lagged saving rate, i.e., $y_{i,t-1}$. Their dataset includes 56 countries over the period of 1995 – 2010. Hence, we have a dynamic panel data model with $N = 56$, $T = 15$, and $k = 5$ (including the constant).

Table 6 shows the values of AIC, BIC, BIC_2 and CV. AIC, BIC_2 and CV all select Model 2, while BIC selects Model 1. Considering the poor performance of BIC in the simulations, we conclude that Model 2 (i.e., including individual fixed effects only) is the correct model.

4.3 Application III: Guns and crime in the U.S.

In the paper "Shooting down the 'More Guns less Crime' hypothesis", Ayres and Donohue (2003) consider how the "shall-issue" law affects the crime rates in the U.S., where the "shall-issue" law refers to whether local authorities can issue a concealed weapon permit if the applicants meet certain determinate criteria. So, here y_{it} is the crime rates for state i in year t. Specifically, we consider the logarithms of three measures of crime rates separately, namely, the violent crime rate, the murder rate and the robbery rate, which are measured by incidents per 100,000 members of the population. The key regressor in x_{it} is the "shall-issue" variable, which is 1 if the state has a shall-carry law in effect in that year and 0 otherwise. Other controls in x_{it} include (*i*) the incarceration rate in the state in the previous year, which is measured by sentenced prisoners per 100,000 residents, (*ii*) the population density per square mile of land area, divided by 1000, (*iii*) the real per capita personal income in the state, in thousands of dollars, (*iv*) the state population, in millions of people, (*v*) the percentage of state population that is male with an age between 10 and 29, (*vi*) the percentage of state population that is white with an age between 10 to 64 and (*vii*) the percentage of state population that is black with an age between 10 and 64. The dataset contains 50 US states and the District of Columbia ($N = 51$) over the period of 1977 – 1999 ($T = 23$). The dataset is also discussed in the textbook by Stock and Watson (2012).

We first consider a static panel model, where the dimension of x_{it} is $k = 9$ (including constant).

Table 7 shows the results for three dependent variables separately. All the information criteria and CV methods select Model 4 (i.e., including both individual and time fixed effects).

We then consider a dynamic panel model by including the lagged y_{it} as an additional regressor, hence here $k = 10$ and $(N, T) = (51, 22)$. Table 7 presents the values of AIC, BIC, BIC_2 and CV for the three dependent variables. AIC, BIC_2 and CV all select Model 4 for the three cases. When y_{it} is the violent crime rate or the robbery rate, BIC chooses Model 3. When y_{it} is the murder rate, BIC chooses Model 1. Given the poor performance of BIC in the simulations, we conclude that for the dynamic panel model, the correct model is also Model 4.

5 Conclusion

In this paper, we propose a jackknife method to determine fixed effects in panel data models based on the leave-one-out cross validation (CV) criterion function. We show that when the serial correlation and cross-sectional dependence in the error terms are weak, our new method can consistently select the correct model. We also modify the CV criterion function to take into account the strong serial correlation in the error term. Our simulations suggest that our new method outperforms the methods based on the information criteria such as AIC and BIC. We provide three empirical applications on (*i*) the crime rates in North Carolina, (*ii*) the determinants of saving rates across countries, and (*iii*) the relationship between guns and crime rates in the U.S.

Several extension are possible. First, our method can be extended to multidimensional panel data models where there are many ways of incorporating fixed effects (see, e.g., Balazsi, Matyas, and Wansbeek (2016) for a review). Therefore, it is even more imperative to select an appropriate specification of fixed effects in multidimensional panels. Second, we may extend our method to allow for strong form of cross-sectional dependence, say, via the multi-factor error structure (e.g., Bai (2009) and Pesaran (2006)). When the regressors also share the factor structure, we conjecture that we can augment Models 1-4 by the cross-sectional means of the dependent and independent variables and then apply our jackknife method. We shall explore these topics in our future research.

Appendix

A Proofs of the main results

To prove Theorem 2.1, we first state and prove six lemmas.

Lemma A.1 *Let $X_D = (X, D)$ and $M_D = I_{NT} - D(D'D)^{-1} D'$. If both $D'D$ and $X'M_DX$ are nonsingular, then*

$$(X_D'X_D)^{-1} = \begin{pmatrix} X_D^* & -X_D^*X'D(D'D)^{-1} \\ -(D'D)^{-1} D'XX_D^* & (D'D)^{-1} + (D'D)^{-1} D'XX_D^*X'D(D'D)^{-1} \end{pmatrix}$$

where $X_D^* = (X'M_DX)^{-1}$.

Proof. Noting that $X_D'X_D = \begin{pmatrix} X'X & X'D \\ D'X & D'D \end{pmatrix}$, the lemma follows from the standard inversion formula for a 2×2 partitioned matrix. See, e.g., Bernstein (2005, p.45). ■

Lemma A.2 *Let $X_D = (X, D)$ and $D = (D_1, D_2)$ where $D_1'D_2 = 0$. If $D_1'D_1$, $D_2'D_2$, and $X'M_DX$ are all nonsingular, then*

$$(X_D'X_D)^{-1} = \begin{pmatrix} X_D^* & -X_D^*B_1 & -X_D^*B_2 \\ -B_1'X_D^* & (D_1'D_1)^{-1} + B_1'X_D^*B_1 & B_1'X_D^*B_2 \\ -B_2'X_D^* & B_2'X_D^*B_1 & (D_2'D_2)^{-1} + B_2'X_D^*B_2 \end{pmatrix}$$

where $X_D^* = (X'M_DX)^{-1}$ and $B_l = X'D_l(D_l'D_l)^{-1}$ for $l = 1, 2$.

Proof. By Lemma A.1,

$$(X_D'X_D)^{-1} = \begin{pmatrix} X_D^* & -X_D^*X'D(D'D)^{-1} \\ -(D'D)^{-1} D'XX_D^* & (D'D)^{-1} + (D'D)^{-1} D'XX_D^*X'D(D'D)^{-1} \end{pmatrix}.$$

Noting that $D_1'D_2 = 0$, we have

$$\begin{aligned} (D'D)^{-1} &= \begin{pmatrix} (D_1'D_1)^{-1} & \\ & (D_2'D_2)^{-1} \end{pmatrix}, \text{ and} \\ X'D(D'D)^{-1} &= X'\left(D_1(D_1'D_1)^{-1}, D_2(D_2'D_2)^{-1}\right) = (B_1, B_2). \end{aligned}$$

Combining the above results yields the desired result. ■

Lemma A.3 *Suppose that Assumption A.1(iv) holds. Then*
(i) $\frac{1}{NT}U'D_\alpha(D_\alpha'D_\alpha)^{-1} D_\alpha'U = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 - \bar{u}_{\cdot\cdot}^2 = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 - O_P((NT)^{-1})$,
(ii) $\frac{1}{NT}U'D_\lambda(D_\lambda'D_\lambda)^{-1} D_\lambda'U = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 - \bar{u}_{\cdot\cdot}^2 = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 - O_P((NT)^{-1})$,
(iii) $\frac{1}{NT}U'D_{\alpha\lambda}(D_{\alpha\lambda}'D_{\alpha\lambda})^{-1} D_{\alpha\lambda}'U = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 - 2\bar{u}_{\cdot\cdot}^2 = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 - O_P((NT)^{-1})$.

Proof. (i) Noting that $D_\alpha'D_\alpha = T\left(I_{N-1} + \iota_{N-1}\iota_{N-1}'\right)$, we have

$$(D_\alpha'D_\alpha)^{-1} = T^{-1}(I_{N-1} - \frac{1}{N}\iota_{N-1}\iota_{N-1}'), \tag{A.1}$$

and

$$
\begin{aligned}
D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' &= T^{-1}\left(\left(\begin{array}{c} I_{N-1} \\ -\iota_{N-1}' \end{array}\right)\otimes \iota_T\right)\left(I_{N-1}-\frac{1}{N}\iota_{N-1}\iota_{N-1}'\right)\left(\left(\begin{array}{cc} I_{N-1} & -\iota_{N-1} \end{array}\right)\otimes \iota_T'\right) \\
&= T^{-1}\left(\left(\begin{array}{c} I_{N-1}-\frac{1}{N}\iota_{N-1}\iota_{N-1}' \\ -\frac{1}{N}\iota_{N-1}' \end{array}\right)\otimes \iota_T\right)\left(\left(\begin{array}{cc} I_{N-1} & -\iota_{N-1} \end{array}\right)\otimes \iota_T'\right) \\
&= \left(\begin{array}{cc} I_{N-1}-\frac{1}{N}\iota_{N-1}\iota_{N-1}' & -\frac{1}{N}\iota_{N-1} \\ -\frac{1}{N}\iota_{N-1}' & \frac{N-1}{N} \end{array}\right)\otimes\left(\iota_T\iota_T'/T\right).
\end{aligned}
$$

By straightforward but tedious algebra we can show that

$$
\begin{aligned}
& \frac{1}{NT}U' D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' U \\
= & \frac{1}{NT^2}\left\{\underline{u}_{N-1}'\left[\left(I_{N-1}-\frac{1}{N}\iota_{N-1}\iota_{N-1}'\right)\otimes\left(\iota_T\iota_T'\right)\right]\underline{u}_{N-1}-\frac{2}{N}\underline{u}_{N-1}'\left[\iota_{N-1}\otimes\left(\iota_T\iota_T'\right)\right]u_N+\frac{N-1}{N}u_N'\iota_T\iota_T'u_N\right\} \\
= & \left[\frac{1}{N}\sum_{i=1}^{N-1}\bar{u}_{i\cdot}^2-\left(\frac{1}{N}\sum_{i=1}^{N-1}\bar{u}_{i\cdot}^2\right)^2\right]-\frac{2}{N^2}\sum_{i=1}^{N-1}\bar{u}_{i\cdot}\bar{u}_{N\cdot}+\frac{N-1}{N^2}\bar{u}_{N\cdot}^2 \\
= & \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2-\bar{u}_{\cdot\cdot}^2=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2-O_P((NT)^{-1}),
\end{aligned}
$$

where we use the fact that $\bar{u}_{\cdot\cdot}=\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T u_{it}=O_P\left((NT)^{-1/2}\right)$ by Assumption A.1(iv).

(ii) The proof is analogous to that of (i) and thus omitted. The main difference is that one now applies

$$
D_\lambda\left(D_\lambda' D_\lambda\right)^{-1}D_\lambda'=\left(\iota_N\iota_N'/N\right)\otimes\left(\begin{array}{cc} I_{T-1}-\frac{1}{N}\iota_{T-1}\iota_{T-1}' & -\frac{1}{N}\iota_{T-1} \\ -\frac{1}{N}\iota_{T-1}' & \frac{T-1}{T}\end{array}\right).
$$

(iii) Noting that $D_{\alpha\lambda}\left(D_{\alpha\lambda}' D_{\alpha\lambda}\right)^{-1}D_{\alpha\lambda}'=D_\alpha\left(D_\alpha' D_\alpha\right)^{-1}D_\alpha'+D_\lambda\left(D_\lambda' D_\lambda\right)^{-1}D_\lambda'$ by the fact $D_\alpha' D_\lambda=0$, we have

$$
\begin{aligned}
\frac{1}{NT}U' D_{\alpha\lambda}\left(D_{\alpha\lambda}' D_{\alpha\lambda}\right)^{-1}D_{\alpha\lambda}' U &= \frac{1}{NT}U' D_\alpha\left(D_\alpha' D_\alpha\right)^{-1}D_\alpha' U+\frac{1}{NT}U' D_\lambda\left(D_\lambda' D_\lambda\right)^{-1}D_\lambda' U \\
&= \left(\frac{1}{N}\sum_{i=1}^N\bar{u}_{i\cdot}^2-\bar{u}_{\cdot\cdot}^2\right)+\left(\frac{1}{T}\sum_{t=1}^T\bar{u}_{\cdot t}^2-\bar{u}_{\cdot\cdot}^2\right) \\
&= \frac{1}{N}\sum_{i=1}^N\bar{u}_{i\cdot}^2+\frac{1}{T}\sum_{t=1}^T\bar{u}_{\cdot t}^2-O_P((NT)^{-1}).
\end{aligned}
$$

where the second equality follows from the results in (i)-(ii) and the last equality follows by Assumption A.1(iv). ■

Lemma A.4 *Suppose that Assumptions A.1(iv) and A.2(iii)-(iv) holds. Then*

(i) $\frac{1}{NT}X' D_\alpha\left(D_\alpha' D_\alpha\right)^{-1}D_\alpha' U=\frac{1}{N}\sum_{i=1}^N\bar{x}_{i\cdot}\bar{u}_{i\cdot}-\bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}=O_P(T^{-1}+(NT)^{-1/2}),$

(ii) $\frac{1}{NT}X' D_\lambda\left(D_\lambda' D_\lambda\right)^{-1}D_\lambda' U=\frac{1}{T}\sum_{t=1}^T\bar{x}_{t\cdot}\bar{u}_{t\cdot}-\bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}=O_P(N^{-1}+(NT)^{-1/2}),$

(iii) $\frac{1}{NT}X' D_{\alpha\lambda}\left(D_{\alpha\lambda}' D_{\alpha\lambda}\right)^{-1}D_{\alpha\lambda}' U=\frac{1}{N}\sum_{i=1}^N\bar{x}_{i\cdot}\bar{u}_{i\cdot}+\frac{1}{T}\sum_{t=1}^T\bar{x}_{t\cdot}\bar{u}_{t\cdot}-2\bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}=O_P(N^{-1}+T^{-1}).$

Proof. (i) Following the proof of Lemma A.3(i), we have

$$
\begin{aligned}
& \frac{1}{NT} X' D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' U \\
= & \frac{1}{NT}\left\{ \underline{x}_{N-1}' \left[\left(I_{N-1} - \frac{1}{N}\iota_{N-1}\iota_{N-1}'\right) \otimes (\iota_T \iota_T')\right] \underline{u}_{N-1} - \frac{1}{N} x_{N-1}' \left[\iota_{N-1}' \otimes (\iota_T\iota_T')\right] \underline{u}_{N-1} \right. \\
& \left. -\frac{1}{N} \underline{x}_{N-1}' \left[\iota_{N-1} \otimes (\iota_T \iota_T')\right] u_N + \frac{N-1}{N} x_N' \iota_T \iota_T' u_N \right\} \\
= & \left[\frac{1}{N}\sum_{i=1}^{N-1} \bar{x}_{i\cdot}\bar{u}_{i\cdot} - \frac{1}{N^2}\sum_{i=1}^{N-1}\bar{x}_{i\cdot}\sum_{i=1}^{N-1}\bar{u}_{i\cdot}\right] - \frac{1}{N^2}\bar{x}_{N\cdot}\sum_{i=1}^{N-1}\bar{u}_{i\cdot} - \frac{1}{N^2}\sum_{i=1}^{N-1}\bar{x}_{i\cdot}\bar{u}_{N\cdot} + \frac{N-1}{N^2}\bar{x}_{N\cdot}\bar{u}_{N\cdot} \\
= & \frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot} - \bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot} - O_P((NT)^{-1/2}) = O_P(T^{-1} + (NT)^{-1/2}),
\end{aligned}
$$

where we use the fact that $\frac{1}{N}\sum_{i=1}^N \bar{x}_{i\cdot}\bar{u}_{i\cdot} = \frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T x_{it}u_{is} = O_P\left(T^{-1} + (NT)^{-1/2}\right)$ and $\bar{u}_{\cdot\cdot} = O_P\left((NT)^{-1/2}\right)$ by Assumptions A.2(iii) and A.1(iv).

(ii) The proof is analogous to that of (i) and thus omitted.

(iii) Noting that $D_{\alpha\lambda}\left(D_{\alpha\lambda}' D_{\alpha\lambda}\right)^{-1} D_{\alpha\lambda}' = D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' + D_\lambda \left(D_\lambda' D_\lambda\right)^{-1} D_\lambda'$, the results follow from (i)-(ii) and the fact that $N^{-1} + T^{-1} \geq 2(NT)^{-1/2}$. ■

Lemma A.5 *Let* $\eta_{it}^{(l)} = x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U$ *and* $J_{lNT} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\eta_{it}^{(l)})^2 = \frac{1}{NT}U'X^{(l)}\left(X^{(l)\prime}X^{(l)}\right)^{-1}$ $\times X^{(l)\prime}U$ *for* $l = 1,2,3,4$. *Suppose that Assumptions A.1(iv)-(v) and A.2(iii)-(iv) hold. Then*

(i) $J_{1NT} = O_P((NT)^{-1})$,

(ii) $J_{2NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P((NT)^{-1} + T^{-2})$,

(iii) $J_{3NT} = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P((NT)^{-1} + N^{-2})$,

(iv) $J_{4NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(T^{-2} + N^{-2}\right)$, $J_{4NT} - J_{2NT} = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(N^{-2} + (NT)^{-1}\right)$, *and* $J_{4NT} - J_{3NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P\left(T^{-2} + (NT)^{-1}\right)$.

Proof. (i) $J_{1NT} \leq \left\|\left(\frac{1}{NT}X'X\right)^{-1}\right\| \left\|\frac{1}{NT}X'U\right\|^2 = O_P\left(\frac{1}{NT}\right)$ by Assumption A.1(iv)-(v).

(ii) By Lemma A.1 with $D = D_\alpha$, we have

$$
\begin{aligned}
J_{2NT} &= \frac{1}{NT}U'X^{(2)}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U \\
&= \frac{1}{NT}\left(U'X, U'D_\alpha\right)\begin{pmatrix} X_{D_\alpha}^* & -X_{D_\alpha}^* B_\alpha \\ -B_\alpha' X_{D_\alpha}^* & \left(D_\alpha' D_\alpha\right)^{-1} + B_\alpha' X_D^* B_\alpha \end{pmatrix}\begin{pmatrix} X'U \\ D_\alpha' U\end{pmatrix} \\
&= \frac{1}{NT}\left(U'XX_{D_\alpha}^* - U'D_\alpha B_\alpha' X_{D_\alpha}^*, -U'XX_{D_\alpha}^* B_\alpha + U'D_\alpha\left(D_\alpha' D_\alpha\right)^{-1} + U'D_\alpha B_\alpha' X_{D_\alpha}^* B_\alpha\right)\begin{pmatrix} X'U \\ D_\alpha' U\end{pmatrix} \\
&= \frac{1}{NT}\left(U'XX_{D_\alpha}^* X'U - 2U'D_\alpha B_\alpha' X_{D_\alpha}^* X'U + U'D_\alpha\left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' U + U'D_\alpha B_\alpha' X_{D_\alpha}^* B_\alpha D_\alpha' U\right) \\
&\equiv J_{2NT,1} - 2J_{2NT,2} + J_{2NT,3} + J_{2NT,4}, \text{ say,}
\end{aligned}
$$

where $B_\alpha = X'D_\alpha\left(D_\alpha' D_\alpha\right)^{-1}$. As in (i), we can show that $J_{2NT,1} = O_P((NT)^{-1})$ by Assumption A.1(iv)-

(v). By Lemma A.4(i) and Assumptions A.1(iv)-(v) and A.2(iii) and using $X^*_{D_\alpha} = (X'M_{D_\alpha}X)^{-1}$,

$$\begin{aligned}
J_{2NT,1} &= \frac{1}{NT}U'D_\alpha B'_\alpha X^*_{D_\alpha} X'U \\
&= \frac{1}{NT}U'D_\alpha (D'_\alpha D_\alpha)^{-1} D'_\alpha X \left(\frac{1}{NT}X'M_{D_\alpha}X\right)^{-1} \frac{1}{NT}X'U \\
&= \left(\frac{1}{N}\sum_{i=1}^N \bar{x}_{i\cdot}\bar{u}_{i\cdot} - \bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}\right)' O_P(1)\, O_P\left((NT)^{-1/2}\right) \\
&= O_P\left(T^{-1} + (NT)^{-1/2}\right) O_P\left((NT)^{-1/2}\right) = O_P\left(N^{-1/2}T^{-3/2} + (NT)^{-1}\right),
\end{aligned}$$

and

$$\begin{aligned}
J_{2NT,4} &= \frac{1}{NT}U'D_\alpha B'_\alpha X^*_{D_\alpha} B_\alpha D'_\alpha U \\
&= \frac{1}{NT}U'D_\alpha (D'_\alpha D_\alpha)^{-1} D'_\alpha X \left(\frac{1}{NT}X'M_{D_\alpha}X\right)^{-1} \frac{1}{NT}X'D_\alpha (D'_\alpha D_\alpha)^{-1} D'_\alpha U \\
&= \left(\frac{1}{N}\sum_{i=1}^N \bar{x}_{i\cdot}\bar{u}_{i\cdot} - \bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}\right)' \left(\frac{1}{NT}X'M_{D_\alpha}X\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^N \bar{x}_{i\cdot}\bar{u}_{i\cdot} - \bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot}\right) \\
&= O_P\left(T^{-1} + (NT)^{-1/2}\right) O_P(1)\, O_P\left(T^{-1} + (NT)^{-1/2}\right) = O_P\left(T^{-2} + (NT)^{-1}\right).
\end{aligned}$$

By Lemma A.3(i),

$$J_{2NT,3} = \frac{1}{NT}U'D_\alpha (D'_\alpha D_\alpha)^{-1} D'_\alpha U = \frac{1}{N}\sum_{i=1}^N \bar{u}^2_{i\cdot} - \bar{u}^2_{\cdot\cdot} = \frac{1}{N}\sum_{i=1}^N \bar{u}^2_{i\cdot} - O_P((NT)^{-1}).$$

It follows that $J_{2NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}^2_{i\cdot} + O_P((NT)^{-1} + T^{-2})$ as $(NT)^{-1} + T^{-2} \geq 2N^{-1/2}T^{-3/2}$.

(iii) The proof is analogous to that of (ii) and thus omitted.

(iv) By Lemma A.2

$$\begin{aligned}
J_{4NT} &= \frac{1}{NT}U'X^{(4)}\left(X^{(4)\prime}X^{(4)}\right)^{-1} X^{(4)\prime}U \\
&= \frac{1}{NT}(U'X, U'D_\alpha, U'D_\lambda) \\
&\quad \times \begin{pmatrix} X^*_{D_{\alpha\lambda}} & -X^*_{D_{\alpha\lambda}}B_\alpha & -X^*_{D_{\alpha\lambda}}B_\lambda \\ -B'_\alpha X^*_{D_{\alpha\lambda}} & (D'_\alpha D_\alpha)^{-1} + B'_\alpha X^*_{D_{\alpha\lambda}}B_\alpha & B'_\alpha X^*_{D_{\alpha\lambda}}B_\lambda \\ -B'_\lambda X^*_{D_{\alpha\lambda}} & B'_\lambda X^*_{D_{\alpha\lambda}}B_\alpha & (D'_\lambda D_\lambda)^{-1} + B'_\lambda X^*_{D_{\alpha\lambda}}B_\lambda \end{pmatrix} \begin{pmatrix} X'U \\ D'_\alpha U \\ D'_\lambda U \end{pmatrix} \\
&= \frac{1}{NT}\Big\{U'XX^*_{D_{\alpha\lambda}}X'U + U'D_\alpha\left((D'_\alpha D_\alpha)^{-1} + B'_\alpha X^*_{D_{\alpha\lambda}}B_\alpha\right)D'_\alpha U \\
&\quad + U'D_\lambda\left((D'_\lambda D_\lambda)^{-1} + B'_\lambda X^*_{D_{\alpha\lambda}}B_\lambda\right)D'_\lambda U - 2U'D_\alpha B'_\alpha X^*_{D_{\alpha\lambda}}X'U \\
&\quad - 2U'D_\lambda B'_\lambda X^*_{D_{\alpha\lambda}}X'U + 2U'D_\lambda B'_\lambda X^*_{D_{\alpha\lambda}}B_\alpha D'_\alpha U\Big\} \\
&\equiv J_{4NT,1} + J_{4NT,2} + J_{4NT,3} - 2J_{4NT,4} - 2J_{4NT,5} + 2J_{4NT,6}, \text{ say,}
\end{aligned}$$

where $D_{\alpha\lambda} = (D_\alpha, D_\lambda)$ and $B_\ell = X'D_\ell (D'_\ell D_\ell)^{-1}$ for $\ell = \alpha, \lambda$. By Assumption A.1(iv)-(v),

$$J_{4NT,1} \leq \left\|\left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1}\right\| \left\|\frac{1}{NT}X'U\right\|^2 = O_P((NT)^{-1}).$$

By Lemmas A.3(i) and A.4(i) and Assumptions A.1(iv) and A.2(iii),

$$\begin{aligned} J_{4NT,2} &= \frac{1}{NT}U'D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' U + \frac{1}{NT}U'D_\alpha B_\alpha' X^*_{D_{\alpha\lambda}} B_\alpha D_\alpha' U \\ &= \frac{1}{NT}U'D_\alpha \left(D_\alpha' D_\alpha\right)^{-1} D_\alpha' U + \frac{1}{NT}U'D_\alpha B_\alpha' \left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1} \frac{1}{NT}B_\alpha D_\alpha' U \\ &= \left(\frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 - \bar{u}_{\cdot\cdot}^2\right) + O_P\left(T^{-1} + (NT)^{-1/2}\right) O_P(1)\, O_P\left(T^{-1} + (NT)^{-1/2}\right) \\ &= \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P\left(T^{-2} + (NT)^{-1}\right). \end{aligned}$$

By Lemmas A.3(ii) and A.4(ii) and Assumptions A.1(iv) and A.2(iv),

$$\begin{aligned} J_{4NT,3} &= \frac{1}{NT}U'D_\lambda \left(D_\lambda' D_\lambda\right)^{-1} D_\lambda' U + \frac{1}{NT}U'D_\lambda B_\lambda' \left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1} \frac{1}{NT}B_\lambda D_\lambda' U \\ &= \left(\frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 - \bar{u}_{\cdot\cdot}^2\right) + O_P\left(N^{-1} + (NT)^{-1/2}\right) O_P(1)\, O_P\left(N^{-1} + (NT)^{-1/2}\right) \\ &= \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(N^{-2} + (NT)^{-1}\right). \end{aligned}$$

In addition, by Lemma A.4(i)-(ii) and Assumption A.1(iv), we have

$$\begin{aligned} J_{4NT,4} &= \frac{1}{NT}U'D_\alpha B_\alpha' \left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1} \frac{1}{NT}X'U \\ &= O_P(T^{-1} + (NT)^{-1/2}) O_P(1)\, O_P((NT)^{-1/2}) = O_P((NT)^{-1} + N^{-1/2}T^{-3/2}), \\ J_{4NT,5} &= \frac{1}{NT}U'D_\lambda B_\lambda' \left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1} \frac{1}{NT}X'U \\ &= O_P(N^{-1} + (NT)^{-1/2}) O_P(1)\, O_P((NT)^{-1/2}) = O_P((NT)^{-1} + N^{-3/2}T^{-1/2}), \\ J_{4NT,6} &= \frac{1}{NT}U'D_\lambda B_\lambda' \left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)^{-1} \frac{1}{NT}B_\alpha D_\alpha' U \\ &= O_P(N^{-1} + (NT)^{-1/2}) O_P(1)\, O_P(T^{-1} + (NT)^{-1/2}) = O_P((NT)^{-1} + N^{-3/2}T^{-1/2} + N^{-1/2}T^{-3/2}). \end{aligned}$$

It follows that $J_{4NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(T^{-2} + N^{-2}\right)$. In addition, we can show that $J_{4NT} - J_{2NT} = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P(N^{-2} + (NT)^{-1})$ and $J_{4NT} - J_{3NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P(T^{-2} + (NT)^{-1})$. ■

Lemma A.6 *Let* $h_{it}^{(l)} = x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} x_{it}^{(l)}$ *for* $l = 1, 2, 3, 4$ *and* $B_\ell = X'D_\ell \left(D_\ell' D_\ell\right)^{-1}$ *for* $\ell = \alpha, \lambda,$ *and* $\alpha\lambda$. *Let* $\max_{i,t} = \max_{1\leq i\leq N, 1\leq t\leq T}$. *Suppose that Assumption A.1(ii) and (v) hold. Then*

(i) $\max_{i,t} h_{it}^{(1)} = o_P((NT)^{-1} C_{NT}^2)$,

(ii) $h_{it}^{(2)} = T^{-1}\frac{N-1}{N} + (x_{it} - B_\alpha d_{\alpha,it})' X^*_{D_\alpha}(x_{it} - B_\alpha d_{\alpha,it})$ *and* $\max_{i,t} h_{it}^{(2)} = o_P((NT)^{-1} C_{NT}^2) + O_P(T^{-1})$,

(iii) $h_{it}^{(3)} = N^{-1}\frac{T-1}{T} + (x_{it} - B_\lambda d_{\lambda,it})' X^*_{D_\lambda}(x_{it} - B_\lambda d_{\lambda,it})$ *and* $\max_{i,t} h_{it}^{(3)} = o_P((NT)^{-1} C_{NT}^2) + O_P(N^{-1})$,

(iv) $h_{it}^{(4)} = T^{-1}\frac{N-1}{N} + N^{-1}\frac{T-1}{T} + (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})' X^*_{D_{\alpha\lambda}}(x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})$ *and* $\max_{i,t} h_{it}^{(4)} = o_P((NT)^{-1} C_{NT}^2) + O_P(N^{-1} + T^{-1})$.

Proof. (i) $\max_{i,t} h_{it}^{(1)} = \max_{i,t} x_{it}' (X'X)^{-1} x_{it} \leq \left[\lambda_{\min}\left((NT)^{-1} X'X\right)\right]^{-1} (NT)^{-1} \max_{i,t} \|x_{it}\|^2 = o_P((NT)^{-1} C_{NT}^2)$ by Assumption A.1(ii) and (v).

(ii) Let $d_{\alpha,it}'$ denote a typical row of D_α such that $D_\alpha = (d_{\alpha,11}, ..., d_{\alpha,1T}, ..., d_{\alpha,N1}, ..., d_{\alpha,NT})'$. Then

$$\begin{aligned}
h_{it}^{(2)} &= x_{it}^{(2)\prime} \left(X^{(2)\prime} X^{(2)}\right)^{-1} x_{it}^{(2)} \\
&= (x_{it}', d_{\alpha,it}') \begin{pmatrix} X'X & X'D_\alpha \\ D_\alpha' X & D_\alpha' D_\alpha \end{pmatrix}^{-1} \begin{pmatrix} x_{it} \\ d_{\alpha,it} \end{pmatrix} \\
&= (x_{it}', d_{\alpha,it}') \begin{pmatrix} X_{D_\alpha}^* & -X_{D_\alpha}^* B_\alpha \\ -B_\alpha' X_{D_\alpha}^* & (D_\alpha' D_\alpha)^{-1} + B_\alpha' X_{D_\alpha}^* B_\alpha \end{pmatrix} \begin{pmatrix} x_{it} \\ d_{\alpha,it} \end{pmatrix} \\
&= d_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} d_{\alpha,it} + x_{it}' X_{D_\alpha}^* x_{it} - d_{\alpha,it}' B_\alpha' X_{D_\alpha}^* x_{it} - x_{it}' X_{D_\alpha}^* B_\alpha d_{\alpha,it} + d_{\alpha,it}' B_\alpha' X_{D_\alpha}^* B_\alpha d_{\alpha,it} \\
&= d_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} d_{\alpha,it} + (x_{it} - B_\alpha d_{\alpha,it})' X_{D_\alpha}^* (x_{it} - B_\alpha d_{\alpha,it}). \qquad \text{(A.2)}
\end{aligned}$$

For $i \leq N-1$, $d_{\alpha,it}$ contains 1 in one place and zeros elsewhere, implying that $d_{\alpha,it}' \left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}'\right) d_{\alpha,it} = 1 - \frac{1}{N} = \frac{N-1}{N}$ for any $i \leq N-1$ and $t = 1, ..., T$. When $i = N$, we have

$$d_{\alpha,Nt}' \left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}'\right) d_{\alpha,Nt} = \iota_{N-1}' \left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}'\right) \iota_{N-1} = \frac{N-1}{N} \text{ for } t = 1, ..., T.$$

These observations, in conjunction with (A.1), imply that

$$d_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} d_{\alpha,it} = T^{-1} d_{\alpha,it}' \left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}'\right) d_{\alpha,it} = T^{-1} \frac{N-1}{N} \text{ for all } i, t. \qquad \text{(A.3)}$$

Next, notice that

$$\begin{aligned}
\max_{i,t} (x_{it} - B_\alpha d_{\alpha,it})' X_{D_\alpha}^* (x_{it} - B_\alpha d_{\alpha,it}) &\leq \epsilon_{NT} \max_{i,t} \frac{1}{NT} (x_{it} - B_\alpha d_{\alpha,it})' (x_{it} - B_\alpha d_{\alpha,it}) \\
&\leq \epsilon_{NT} \max_{i,t} \frac{1}{NT} \|x_{it}\|^2 = o_P((NT)^{-1} C_{NT}^2),
\end{aligned}$$

where $\epsilon_{NT} = \left[\lambda_{\min}\left(\frac{1}{NT} X' M_{D_\alpha} X\right)\right]^{-1} = O_P(1)$ by Assumption A.1(v) and we use the fact that $x_{it} - B_\alpha d_{\alpha,it}$ denotes the residual from the OLS regression of x_{it} on $d_{\alpha,it}$. It follows that $h_{it}^{(2)} = T^{-1} \frac{N-1}{N} + (x_{it} - B_\alpha d_{\alpha,it})' X_{D_\alpha}^* (x_{it} - B_\alpha d_{\alpha,it})$ and $\max_{i,t} h_{it}^{(2)} = o_P((NT)^{-1} C_{NT}^2) + O_P(T^{-1})$.

(iii) Let $d_{\lambda,it}'$ denote a typical row of D_λ such that $D_\lambda = (d_{\lambda,11}, ..., d_{\lambda,1T}, ..., d_{\lambda,N1}, ..., d_{\lambda,NT})'$. Following the analysis in (ii), we can show that

$$h_{it}^{(3)} = d_{\lambda,it}' (D_\lambda' D_\lambda)^{-1} d_{\lambda,it} + (x_{it} - B_\lambda d_{\lambda,it})' X_{D_\lambda}^* (x_{it} - B_\lambda d_{\lambda,it}) \qquad \text{(A.4)}$$

and

$$d_{\alpha,it}' \left(I_{T-1} - \frac{1}{T} \iota_{T-1} \iota_{T-1}'\right) d_{\alpha,it} = \frac{T-1}{T} \text{ for all } i, t. \qquad \text{(A.5)}$$

Noting that

$$D_\lambda' D_\lambda = N\left(I_{T-1} + \iota_{T-1} \iota_{T-1}'\right) \text{ and } (D_\lambda' D_\lambda)^{-1} = N^{-1} \left(I_{T-1} - \frac{1}{T} \iota_{T-1} \iota_{T-1}'\right), \qquad \text{(A.6)}$$

we have $d_{\lambda,it}' (D_\lambda' D_\lambda)^{-1} d_{\lambda,it} = N^{-1} \frac{T-1}{T}$. In addition,

$$\max_{i,t} (x_{it} - B_\lambda d_{\lambda,it})' X_{D_\alpha}^* (x_{it} - B_\lambda d_{\lambda,it}) \leq \epsilon_{NT} \max_{i,t} \frac{1}{NT} \|x_{it}\|^2 = o_P((NT)^{-1} C_{NT}^2).$$

It follows that $h_{it}^{(3)} = N^{-1}\frac{T-1}{T} + (x_{it} - B_\lambda d_{\lambda,it})' X_{D_\lambda}^* (x_{it} - B_\lambda d_{\lambda,it})$ and $\max_{i,t} h_{it}^{(3)} = o_P((NT)^{-1} C_{NT}^2) + O_P(N^{-1})$.

(iv) Let $d'_{\alpha\lambda,it}$ denote a typical row of $D_{\alpha\lambda}$ such that $D_{\alpha\lambda} = (d_{\alpha\lambda,11}, ..., d_{\alpha\lambda,1T}, ..., d_{\alpha\lambda,N1}, ..., d_{\alpha\lambda,NT})'$. Following the analysis in (ii), we can show that

$$h_{it}^{(4)} = d'_{\alpha\lambda,it} (D'_{\alpha\lambda} D_{\alpha\lambda})^{-1} d_{\alpha\lambda,it} + (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})' X_{D_{\alpha\lambda}}^* (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it}). \tag{A.7}$$

Noting that $D_{\alpha\lambda} = (D_\alpha, D_\lambda)$ and $D'_\alpha D_\lambda = 0$, we have

$$(D'_{\alpha\lambda} D_{\alpha\lambda})^{-1} = \begin{pmatrix} (D'_\alpha D_\alpha)^{-1} & \\ & (D'_\lambda D_\lambda)^{-1} \end{pmatrix}.$$

Then

$$\begin{aligned} d'_{\alpha\lambda,it} (D'_{\alpha\lambda} D_{\alpha\lambda})^{-1} d_{\alpha\lambda,it} &= d'_{\alpha,it} (D'_\alpha D_\alpha)^{-1} d_{\alpha,Nt} + d'_{\lambda,it} (D'_\lambda D_\lambda)^{-1} d_{\lambda,it} \\ &= T^{-1}\frac{N-1}{N} + N^{-1}\frac{T-1}{T} \text{ for all } i,t. \end{aligned} \tag{A.8}$$

In addition, $\max_{i,t} (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})' X_{D_{\alpha\lambda}}^* (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it}) \le \left[\lambda_{\min}\left(\frac{1}{NT} X' M_{D_{\alpha\lambda}} X\right)\right]^{-1} \max_{i,t} \frac{1}{NT} \|x_{it}\|^2$ $= o_P((NT)^{-1} C_{NT}^2)$ by Assumptions A.1(ii) and (v). It follows that $h_{it}^{(4)} = T^{-1}\frac{N-1}{N} + N^{-1}\frac{T-1}{T} + (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})' X_{D_{\alpha\lambda}}^* (x_{it} - B_{\alpha\lambda} d_{\alpha\lambda,it})$ and $\max_{i,t} h_{it}^{(4)} = o_P((NT)^{-1} C_{NT}^2) + O_P(N^{-1} + T^{-1})$. ■

Proof of Theorem 2.1. Recall that $\hat\beta^{(l)} = \left(X^{(l)\prime} X^{(l)}\right)^{-1} X^{(l)\prime} Y$ and $\hat\beta_{it}^{(l)} = (X^{(l)\prime} X^{(l)} - x_{it}^{(l)} x_{it}^{(l)\prime})^{-1} (X^{(l)\prime} Y - x_{it}^{(l)} y_{it})$. By the updated formula for OLS estimation (e.g., Greene (2008, p.964)), we have for $l = 1, 2, 3, 4$,

$$\begin{aligned} \hat\beta_{it}^{(l)} - \hat\beta^{(l)} &= (X^{(l)\prime} X^{(l)} - x_{it}^{(l)} x_{it}^{(l)\prime})^{-1} (X^{(l)} Y - x_{it}^{(l)} y_{it}) - \hat\beta^{(l)} \\ &= \left[\left(X^{(l)\prime} X^{(l)}\right)^{-1} + \frac{1}{1 - h_{it}^{(l)}} \left(X^{(l)\prime} X^{(l)}\right)^{-1} x_{it}^{(l)} x_{it}^{(l)\prime} \left(X^{(l)\prime} X^{(l)}\right)^{-1}\right] (X^{(l)} Y - x_{it}^{(l)} y_{it}) - \hat\beta^{(l)} \\ &= \frac{-1}{1 - h_{it}^{(l)}} \left(X^{(l)\prime} X^{(l)}\right)^{-1} x_{it}^{(l)} y_{it} + \frac{1}{1 - h_{it}^{(l)}} \left(X^{(l)\prime} X^{(l)}\right)^{-1} x_{it}^{(l)} x_{it}^{(l)\prime} \left(X^{(l)\prime} X^{(l)}\right)^{-1} X^{(l)} Y, \end{aligned} \tag{A.9}$$

where $h_{it}^{(l)} = x_{it}^{(l)\prime} \left(X^{(l)\prime} X^{(l)}\right)^{-1} x_{it}^{(l)}$. Below, we will use $CV_{l,m}$ to denote the $CV(m)$ when the true model is given by Model l where $l, m = 1, 2, 3, 4$. Let $c_{it,l} = (1 - h_{it}^{(l)})^{-1}$and $c_{it,lm} = c_{it,l} c_{it,m}$. By Lemma A.6, for $l, m = 1, 2, 3, 4$ we have

$$\max_{i,t} h_{it}^{(l)} = o_P(\delta_{NT}), \ \ \max_{i,t} |c_{it,l} - 1| = o_P(\delta_{NT}) \text{ and } \max_{i,t} |c_{it,lm} - 1| = o_P(\delta_{NT}), \tag{A.10}$$

where $\delta_{NT} = (NT)^{-1} C_{NT}^2$ and $a \wedge b = \min(a, b)$.

Case 1: Model 1 is the true model. In this case, Models 2-4 are all overfitted and we will show that $P(CV_{1,1} < CV_{1,m}) \to 1$ for $m = 2, 3, 4$. When Model 1 is true, we have

$$y_{it} = \beta' x_{it} + u_{it} = \beta^{(l)\prime} x_{it}^{(l)} + u_{it} \text{ and } \hat\beta^{(l)} - \beta^{(l)} = \left(X^{(l)\prime} X^{(l)}\right)^{-1} X^{(l)\prime} U,$$

where the true values correspond to the coefficients of the dummies $d_{\alpha,it}$ and $d_{\lambda,it}$ for α_i and λ_t in $\beta^{(l)}$,

$l = 2, 3, 4$, are all zero. This, in conjunction with (A.9), implies that for $l = 1, 2, 3, 4$,

$$\begin{aligned}
& x_{it}^{(l)\prime}(\hat{\beta}_{it}^{(l)} - \beta^{(l)}) \\
= \; & x_{it}^{(l)\prime}\left[(\hat{\beta}^{(l)} - \beta^{(l)}) - \frac{1}{1-h_{it}^{(l)}}\left(X^{(l)\prime}X^{(l)}\right)^{-1} x_{it}^{(l)}u_{it} + \frac{1}{1-h_{it}^{(l)}}\left(X^{(l)\prime}X^{(l)}\right)^{-1} x_{it}^{(l)}x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U\right] \\
= \; & x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U - \frac{h_{it}^{(l)}}{1-h_{it}^{(l)}}u_{it} + \frac{h_{it}^{(l)}}{1-h_{it}^{(l)}}x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U \\
= \; & -\frac{h_{it}^{(l)}}{1-h_{it}^{(l)}}u_{it} + \frac{1}{1-h_{it}^{(l)}}x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U,
\end{aligned}$$

and

$$y_{it} - \hat{y}_{it}^{(l)} = u_{it} - x_{it}^{(l)\prime}\left(\hat{\beta}_{it}^{(l)} - \beta^{(l)}\right) = c_{it,l}\left(u_{it} - x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U\right). \tag{A.11}$$

It follows that

$$CV_{1,l} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{y}_{it}^{(l)}\right)^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,l}^2\left(u_{it} - x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1} X^{(l)\prime}U\right)^2.$$

We first study $CV_{1,2} - CV_{1,1}$. We make the following decomposition:

$$\begin{aligned}
& CV_{1,2} - CV_{1,1} \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,2}^2\left(u_{it} - x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1} X^{(2)}U\right)^2 - c_{it,1}^2\left(u_{it} - x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1} X^{(1)\prime}U\right)^2\right] \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,2}^2 - c_{it,1}^2\right)u_{it}^2 \\
& + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,2}^2\left(x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1} X^{(2)}U\right)^2 - c_{it,1}^2\left(x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1} X^{(1)\prime}U\right)^2\right] \\
& - \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,2}^2u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1} X^{(2)}U - c_{it,1}^2u_{it}x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1} X^{(1)\prime}U\right] \\
\equiv \; & A_1 + A_2 - 2A_3, \text{ say.}
\end{aligned}$$

For A_1, we have

$$\begin{aligned}
A_1 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,2}^2 - c_{it,1}^2\right)u_{it}^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,12}^2\left(2 - h_{it}^{(1)} - h_{it}^{(2)}\right)\left(h_{it}^{(2)} - h_{it}^{(1)}\right)u_{it}^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,12}^2\left(2 - h_{it}^{(1)} - h_{it}^{(2)}\right)h_{it}^{(2)}u_{it}^2 - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{it,12}^2\left(2 - h_{it}^{(1)} - h_{it}^{(2)}\right)h_{it}^{(1)}u_{it}^2 \\
&\equiv A_{1,1} - A_{1,2}, \text{ say.}
\end{aligned}$$

For $A_{1,1}$, we make the following decomposition:

$$\begin{aligned}
A_{1,1} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,12}^2\left(2-h_{it}^{(1)}-h_{it}^{(2)}\right)h_{it}^{(2)}u_{it}^2 \\
&= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} h_{it}^{(2)}u_{it}^2+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,12}^2-1\right)h_{it}^{(2)}u_{it}^2-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,12}^2\left(h_{it}^{(1)}+h_{it}^{(2)}\right)h_{it}^{(2)}u_{it}^2 \\
&\equiv A_{1,11}+A_{1,12}-A_{1,3}.
\end{aligned}$$

By Lemma A.6(ii), we can readily show that

$$\begin{aligned}
A_{1,11} &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[d_{\alpha,it}'\left(D_\alpha' D_\alpha\right)^{-1}d_{\alpha,it}+\left(x_{it}-B_\alpha d_{\alpha,it}\right)' X_{D_\alpha}^*\left(x_{it}-B_\alpha d_{\alpha,it}\right)\right]u_{it}^2 \\
&= T^{-1}\frac{N-1}{N}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2+O_P\left((NT)^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}\right\|^2 u_{it}^2 \\
&= 2T^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2+O_P\left((NT)^{-1}\right).
\end{aligned}$$

This result, in conjunction with (A.10) and the dominated convergence theorem (DCT), implies that $A_{1,12}=o_P\left(T^{-1}\right)$ and $A_{1,13}=o_P\left(T^{-1}\right)$. For $A_{1,2}$, we have by (A.10)

$$\begin{aligned}
A_{1,2} &\leq \max_{i,t} c_{it,12}^2(2-h_{it}^{(1)}-h_{it}^{(2)})\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}'\left(X'X\right)^{-1}x_{it}u_{it}^2 \\
&\leq \frac{1}{NT}\max_{i,t} c_{it,12}^2(2-h_{it}^{(1)}-h_{it}^{(2)})\left[\lambda_{\min}\left(\frac{1}{NT}X'X\right)\right]^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}\right\|^2 u_{it}^2=O_P\left((NT)^{-1}\right).
\end{aligned}$$

It follows that $A_1=2T^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2+o_P\left(T^{-1}\right)$. For A_2, we write

$$\begin{aligned}
A_2 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,2}^2-1\right)\left(x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^2 \\
&\quad -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,1}^2\left(x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}U\right)^2 \\
&\equiv A_{2,1}+A_{2,2}-A_{23}, \text{ say.}
\end{aligned}$$

By Lemma A.5(ii), $A_{2,1}=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+O_P((NT)^{-1}+T^{-2}+N^{-1/2}T^{-3/2})$, where the first term is $O_P(T^{-1})$. This result, in conjunction with (A.10) and the DCT, implies that $A_{2,2}=o_P(T^{-1})$. By Lemmas A.5(i) and A.6(i), $A_{2,3}=O_P\left((NT)^{-1}\right)$. It follows that $A_2=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+o_P(T^{-1})$. For A_3, we have

$$\begin{aligned}
A_3 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)}U+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,2}^2-1\right)u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)}U \\
&\quad -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)}U+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(1-c_{it,1}^2\right)u_{it}x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)}U \\
&\equiv A_{3,1}+A_{3,2}-A_{33}+A_{34}, \text{ say.}
\end{aligned}$$

By Lemma A.5(i) and (ii), $A_{3,1}=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^{2}+O_P((NT)^{-1}+T^{-2}+N^{-1/2}T^{-3/2})$ and $A_{3,3}=O_P((NT)^{-1})$. In addition,

$$\begin{aligned}
|A_{3,2}| &\leq \max_{i,t}\left|c_{it,2}^{2}-1\right|\left\|\left(\frac{1}{NT}X^{(2)\prime}X^{(2)}\right)^{-1}\right\|\left\|\frac{1}{NT}X^{(2)}U\right\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}^{(2)}u_{it}\right\| \\
&= o_P\left(\delta_{NT}\right)O_P\left(1\right)O_P((NT)^{-1/2})O_P\left(1\right)=o_P(T^{-1}),\text{ and} \\
|A_{3,4}| &\leq \max_{i,t}\left|c_{it,1}^{2}-1\right|\left\|\left(\frac{1}{NT}X^{(1)\prime}X^{(1)}\right)^{-1}\right\|\left\|\frac{1}{NT}X^{(1)}U\right\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}^{(1)}u_{it}\right\| \\
&= o_P\left(\delta_{NT}\right)O_P\left(1\right)O_P((NT)^{-1/2})O_P\left(1\right)=o_P(T^{-1}).
\end{aligned}$$

So $A_3=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^{2}+o_P(T^{-1})$. Combining the above results, we have

$$T\left[CV_{1,2}-CV_{1,1}\right]=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}-\frac{T}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^{2}+o_P\left(1\right)\xrightarrow{P}2\bar{\sigma}_{u}^{2}-\bar{\sigma}_{u1}^{2}, \tag{A.12}$$

where the convergence holds by Assumptions A.1(i) and A.2(i). Similarly, by using Lemma A.5(iii) and Lemma A.6(i) and (iii), we can show that

$$N\left[CV_{1,3}-CV_{1,1}\right]=2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}-\frac{N}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_P\left(1\right)\xrightarrow{P}2\bar{\sigma}_{u}^{2}-\bar{\sigma}_{u2}^{2}, \tag{A.13}$$

where the convergence holds by Assumptions A.1(i) and A.2(ii).

By using Lemma A.5(iv) and Lemma A.6(i) and (iv), we can show that

$$\begin{aligned}
& CV_{1,4}-CV_{1,1} \\
=\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}\left(u_{it}-x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right)^{2}-c_{it,1}^{2}\left(u_{it}-x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}U\right)^{2}\right] \\
=\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,4}^{2}-c_{it,1}^{2}\right)u_{it}^{2} \\
& +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}\left(x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right)^{2}-c_{it,1}^{2}\left(x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}U\right)^{2}\right] \\
& -\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}u_{it}x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U-c_{it,1}^{2}u_{it}x_{it}^{(1)\prime}\left(X^{(1)\prime}X^{(1)}\right)^{-1}X^{(1)\prime}U\right] \\
\equiv\ & A_4+A_5-2A_6,\text{ say.}
\end{aligned}$$

As in the analysis of A_1, we can apply Lemma A.5(iv) and Lemma A.6(i) and (iv) to show that

$$\begin{aligned}
A_4 &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(h_{it}^{(4)}-h_{it}^{(1)})u_{it}^{2}+o_P\left(N^{-1}+T^{-1}\right) \\
&= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{N-1}{NT}+\frac{T-1}{NT}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)^{\prime}X_{D_{\alpha\lambda}}^{*}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)\right]u_{it}^{2}+o_P\left(N^{-1}+T^{-1}\right) \\
&= \left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}+O_P\left(\left(NT\right)^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|x_{it}\right\|^{2}u_{it}^{2}+o_P\left(N^{-1}+T^{-1}\right) \\
&= \left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}+o_P\left(N^{-1}+T^{-1}\right),
\end{aligned}$$

$$
\begin{aligned}
A_5 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)}U\right)^2 + o_P\left(N^{-1}+T^{-1}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2 + o_P\left(N^{-1}+T^{-1}\right), \text{ and} \\
A_6 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)}U + o_P\left(N^{-1}+T^{-1}\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2 + o_P\left(N^{-1}+T^{-1}\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
(N\wedge T)\left[CV_{1,4}-CV_{1,1}\right] &= (N\wedge T)\left[\left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2 - \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 - \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2\right] + o_P(1) \\
&\xrightarrow{P} 2(1+c)\,\bar{\sigma}_u^2 - \left(\bar{\sigma}_{u1}^2 + c\bar{\sigma}_{u2}^2\right)1\{c_1\geq 1\} - \left(c\bar{\sigma}_{u1}^2+\bar{\sigma}_{u2}^2\right)1\{c_1<1\}, \quad \text{(A.14)}
\end{aligned}
$$

where $c=\lim_{(N,T)\to\infty}\left(\frac{N}{T}\wedge\frac{T}{N}\right)$, $c_1=\lim_{(N,T)\to\infty}\frac{N}{T}$, and the convergence holds by Assumptions A.1(i) and A.2(i)-(ii). Combining (A.12)-(A.14) yields $P\left(CV_{1,1}<CV_{1,m}\right)\to 1$ for $m=2,3,4$ provided $\max\left(\bar{\sigma}_{u1}^2,\bar{\sigma}_{u2}^2\right) < 2\bar{\sigma}_u^2$.

Case 2: Model 2 is the true model. In this case, Models 1 and 2 are underfitted and Model 4 is overfitted and we will show that $P\left(CV_{2,2}<CV_{2,m}\right)\to 1$ for $m=1,3,4$. Let $u_{\alpha,it}=\alpha_i+u_{it}$ and $U_\alpha=(u_{\alpha,11},...,u_{\alpha,1T},...,u_{\alpha,N1},...,u_{\alpha,NT})'$. Note that $U_\alpha = D_\alpha\underline{\alpha}+U$ where $\underline{\alpha}=(\alpha_1,...,\alpha_{N-1})'$. Following the steps to obtain (A.11), we can show that

$$
y_{it}-\hat{y}_{it}^{(1)} = u_{\alpha,it}-x_{it}'\left(\hat{\beta}_{it}^{(1)}-\beta^{(1)}\right) = c_{it,1}\left(u_{\alpha,it}-x_{it}'\left(X'X\right)^{-1}X'U_\alpha\right). \quad \text{(A.15)}
$$

Then

$$
\begin{aligned}
CV_{2,1} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{\alpha,it}-x_{it}'\left(X'X\right)^{-1}X'U_\alpha\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,1}^2-1\right)\left(u_{\alpha,it}-x_{it}'\left(X'X\right)^{-1}X'U_\alpha\right)^2 \\
&\equiv A_7 + A_8, \text{ say.}
\end{aligned}
$$

It is easy to show that by Assumptions A.1(i) and (iv)-(vi) and A.3(i)

$$
\begin{aligned}
A_7 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_i - x_{it}'\left(X'X\right)^{-1}X'D_\alpha\underline{\alpha} + u_{it} - x_{it}'\left(X'X\right)^{-1}X'U\right)^2 \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_i - x_{it}'\left(X'X\right)^{-1}X'D_\alpha\underline{\alpha}\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2 + o_P(1) \xrightarrow{P} c_{\alpha,X}+\bar{\sigma}_u^2.
\end{aligned}
$$

This result, in conjunction with (A.10) and the DCT, implies that $A_8=o_P(1)$. In addition, we can follow the analysis in Case 1 and readily show that $CV_{2,2}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2+o_P(1)\xrightarrow{P}\bar{\sigma}_u^2$. It follows that

$$
CV_{2,1}-CV_{2,2}\xrightarrow{P} c_{\alpha,X}>0. \quad \text{(A.16)}
$$

To study $CV_{2,3}$, we observe that

$$y_{it}-\hat{y}_{it}^{(3)}=u_{\alpha,it}-x_{it}^{(3)\prime}\left(\hat{\beta}_{it}^{(3)}-\beta^{(3)}\right)=c_{it,3}\left(u_{\alpha,it}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}U_{\alpha}\right), \qquad \text{(A.17)}$$

and

$$\begin{aligned} CV_{2,3} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{\alpha,it}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}U_{\alpha}\right)^{2} \\ &\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,3}^{2}-1\right)\left(u_{\alpha,it}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}U_{\alpha}\right)^{2} \\ &\equiv A_{9}+A_{10},\ \text{say.} \end{aligned}$$

By Assumptions A.1(i), A.1(iv) and A.3(i), Lemmas A.4-A.5, and (A.10), we can readily show that

$$\begin{aligned} A_{9} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_{i}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_{\alpha}\underline{\alpha}+u_{it}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}U\right)^{2} \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_{i}-x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_{\alpha}\underline{\alpha}\right)^{2}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}+o_{P}(1)\xrightarrow{P}c_{\alpha,X_{\lambda}}+\bar{\sigma}_{u}^{2}, \end{aligned}$$

and $A_{10}=o_{P}(1)$. It follows that

$$CV_{2,3}-CV_{2,2}\xrightarrow{P}c_{\alpha,X_{\lambda}}>0. \qquad \text{(A.18)}$$

To study $CV_{2,4}$, noting that

$$y_{it}-\hat{y}_{it}^{(4)}=u_{it}-x_{it}^{(4)\prime}\left(\hat{\beta}_{it}^{(4)}-\beta^{(4)}\right)=c_{it,4}\left(u_{it}-x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right), \qquad \text{(A.19)}$$

we have

$$\begin{aligned} & CV_{2,4}-CV_{2,2} \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}\left(u_{it}-x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right)^{2}-c_{it,2}^{2}\left(u_{it}-x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^{2}\right] \\ =\ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,4}^{2}-c_{it,2}^{2}\right)u_{it}^{2} \\ & +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}\left(x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right)^{2}-c_{it,2}^{2}\left(x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^{2}\right] \\ & -\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}u_{it}x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U-c_{it,2}^{2}u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right] \\ \equiv\ & A_{11}+A_{12}-2A_{13},\ \text{say.} \end{aligned}$$

Following the analysis of $CV_{1,4}-CV_{1,1}$ in Case 1 and applying Lemmas A.5(ii) and (iv) and A.6 and (A.10), we can readily show that

$$
\begin{aligned}
A_{11} &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_{it}^{(4)}-h_{it}^{(2)}\right)u_{it}^{2}+o_{P}\left(N^{-1}\right)\\
&= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{T-1}{NT}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)'X_{D_{\alpha\lambda}}^{*}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)-\left(x_{it}-B_{\alpha}d_{\alpha,it}\right)'X_{D_{\alpha}}^{*}\left(x_{it}-B_{\alpha}d_{\alpha,it}\right)\right]u_{it}^{2}\\
&\quad +o_{P}\left(N^{-1}\right)\\
&= N^{-1}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}+o_{P}\left(N^{-1}\right),
\end{aligned}
$$

$$
\begin{aligned}
A_{12} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)}U\right)^{2}-\left(x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^{2}\right]+o_{P}\left(N^{-1}\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_{P}\left(N^{-1}\right),\ \text{and}\\
A_{13} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[u_{it}x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)}U-u_{it}x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)}U\right]+o_{P}\left(N^{-1}+T^{-1}\right)\\
&= \frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_{P}\left(N^{-1}\right).
\end{aligned}
$$

It follows that

$$
N\left[CV_{2,4}-CV_{2,2}\right]=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}-\frac{N}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_{P}\left(1\right)\xrightarrow{P}2\bar{\sigma}_{u}^{2}-\bar{\sigma}_{u2}^{2}, \tag{A.20}
$$

where the convergence holds by Assumptions A.1(i) and A.2(ii).

By (A.16), (A.18), and (A.20), we have $P\left(CV_{2,2}<CV_{2,m}\right)\to 1$ as $(N,T)\to\infty$ for $m=1,3,4$ provided $\bar{\sigma}_{u2}^{2}<2\bar{\sigma}_{u}^{2}$.

Case 3: Model 3 is the true model. This case parallels Case 2 and we can analogously show that

$$
\begin{aligned}
CV_{3,1}-CV_{3,3}\xrightarrow{P}c_{\lambda,X} &> 0,\\
CV_{3,2}-CV_{3,3}\xrightarrow{P}c_{\lambda,X_{\alpha}} &> 0,\\
T\left[CV_{3,4}-CV_{3,3}\right]\xrightarrow{P}2\bar{\sigma}_{u}^{2}-\bar{\sigma}_{u1}^{2} &> 0,
\end{aligned}
$$

provided $\bar{\sigma}_{u1}^{2}<2\bar{\sigma}_{u}^{2}$. Then $P\left(CV_{3,3}<CV_{3,m}\right)\to 1$ for $m=1,2,4$.

Case 4: Model 4 is the true model. In this case, Models 1-3 are underfitted and we will show that $P\left(CV_{4,4}<CV_{4,m}\right)\to 1$ for $m=1,2,3$. Let $u_{\lambda,it}=\lambda_{t}+u_{it}$, $u_{\alpha\lambda,it}=\alpha_{i}+\lambda_{t}+u_{it}$, $U_{\lambda}=(u_{\lambda,11},...,u_{\lambda,1T},...,u_{\lambda,N1},...,u_{\lambda,NT})'$, and $U_{\alpha\lambda}=(u_{\alpha\lambda,11},...,u_{\alpha\lambda,1T},...,u_{\alpha\lambda,N1},...,u_{\alpha\lambda,NT})'$. Note that $U_{\alpha\lambda}=D_{\alpha}\underline{\alpha}+D_{\lambda}\underline{\lambda}+U$, where $\underline{\lambda}=(\lambda_{1},...,\lambda_{T-1})'$. Following the steps to obtain (A.11), now we can show that

$$
y_{it}-\hat{y}_{it}^{(1)}=u_{\alpha\lambda,it}-x_{it}'(\hat{\beta}_{it}^{(1)}-\beta^{(1)})=c_{it,1}\left(u_{\alpha\lambda,it}-x_{it}'\left(X'X\right)^{-1}X'U_{\alpha\lambda}\right). \tag{A.21}
$$

As in Case 2, we can show that by Assumptions A.1(i) and (iv)-(vi) and A.3(iii),

$$\begin{aligned} CV_{4,1} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,1}^2 \left[u_{\alpha\lambda,it} - x_{it}'\left(X'X\right)^{-1} X'U_{\alpha\lambda}\right]^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\alpha_i + \lambda_t - x_{it}'\left(X'X\right)^{-1} X'\left(D_\alpha\underline{\alpha} + D_\lambda\underline{\lambda}\right)\right]^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 + o_P(1) \\ &\xrightarrow{P} c_{\alpha\lambda,X} + \bar{\sigma}_u^2. \end{aligned}$$

Similarly, we have by Assumptions A.1(i), A.1(v)-(vi), and A.3(iii)

$$\begin{aligned} CV_{4,2} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,2}^2 \left[u_{\lambda,it} - x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1} X^{(2)\prime}U_\lambda\right]^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\lambda_t - x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1} X^{(2)\prime}D_\lambda\underline{\lambda}\right]^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 + o_P(1) \xrightarrow{P} c_{\lambda,X_\alpha} + \bar{\sigma}_u^2, \\ CV_{4,3} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,3}^2 \left[u_{\alpha,it} - x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1} X^{(3)\prime}U_\alpha\right]^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\alpha_i - x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1} X^{(3)\prime}D_\alpha\underline{\alpha}\right]^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 + o_P(1) \xrightarrow{P} c_{\alpha,X_\lambda} + \bar{\sigma}_u^2, \end{aligned}$$

and

$$CV_{4,4} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,4}\left[u_{it} - x_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1} X^{(4)\prime}U\right]^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 + o_P(1) \xrightarrow{P} \bar{\sigma}_u^2.$$

Then $P\left(CV_{4,4} < CV_{4,m}\right) \to 1$ as $(N,T) \to \infty$ for $m = 1,2,3$. ■

To prove Theorem 2.2, we introduce some notation and prove three lemmas. Let $\hat{\mathbf{u}}_i = (\hat{u}_{i,p+1}, ..., \hat{u}_{i,T,})'$, $\hat{\mathbf{U}} = (\hat{\mathbf{u}}_1, ..., \hat{\mathbf{u}}_N)'$, $\hat{\mathbf{z}}_i = (\underline{\hat{u}}_{i,p}, ..., \underline{\hat{u}}_{i,T-1,})'$, and $\hat{\mathbf{Z}} = (\hat{\mathbf{z}}_1', ..., \hat{\mathbf{z}}_N')'$ where $\underline{\hat{u}}_{i,t} = \underline{\hat{u}}_{i,t}^{(4)} = (\hat{u}_{it}^{(4)}, ..., \hat{u}_{i,t-p+1}^{(4)})'$ for $t = p, ..., T-1$. Let $\mathbf{u}_i = (u_{i,p+1}, ..., u_{i,T,})'$, $\mathbf{U} = (\mathbf{u}_1, ..., \mathbf{u}_N)'$, $z_i = (\underline{\ddot{u}}_{i,p}, ..., \underline{\ddot{u}}_{i,T-1,})'$, and $\mathbf{Z} = (\mathbf{z}_1', ..., \mathbf{z}_N')'$, where $\underline{\ddot{u}}_{i,t} = (\ddot{u}_{it}, ..., \ddot{u}_{i,t-p+1})'$ and $\ddot{u}_{it} = u_{it} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t} + \bar{u}_{\cdot\cdot}$ for $t = p, ..., T-1$. Let $\ddot{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot}$, where $\bar{y}_{i\cdot}$, $\bar{y}_{\cdot t}$, and $\bar{y}_{\cdot\cdot}$ are defined analogously to $\bar{u}_{i\cdot}, \bar{u}_{\cdot t}$, and $\bar{u}_{\cdot\cdot}$.

Lemma A.7 *Suppose Assumptions A.1, A.2 and A.4 hold. Then*
(i) $\frac{1}{NT_p}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}} - \mathbf{Z}'\mathbf{Z}\right) = O_P(\eta_{NT})$,
(ii) $\frac{1}{NT_p}\left(\hat{\mathbf{Z}}'\hat{\mathbf{U}} - \mathbf{Z}'\mathbf{U}\right) = O_P(\eta_{NT})$,
(iii) $\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U} - \boldsymbol{\rho} = O_P(\eta_{NT})$,
where $\eta_{NT} = (NT)^{-1/2} + T^{-1} + N^{-1}$.

Proof. (i) First, we reparametrize Model 4 as

$$y_{it} = x_{it}^{*\prime}\beta^* + \alpha_i^* + \lambda_t + u_{it},$$

where x_{it}^* and β^* correspond to x_{it} and β after one removes the constant term, and α_i^* incorporates the intercept term now. Let $\ddot{x}_{it}^* = x_{it}^* - \bar{x}_{i\cdot}^* - \bar{x}_{\cdot t}^* + \bar{x}_{\cdot\cdot}^*$, where $\bar{x}_{i\cdot}^*$, $\bar{x}_{\cdot t}^*$, and $\bar{x}_{\cdot\cdot}^*$ are defined analogously to $\bar{u}_{i\cdot}, \bar{u}_{\cdot t}$, and $\bar{u}_{\cdot\cdot}$. Let $\ddot{y}_i = (\ddot{y}_{i1}, ..., \ddot{y}_{iT})'$ and $\ddot{Y} = (\ddot{y}_1, ..., \ddot{y}_N)'$. Define $\ddot{x}_i$, $\ddot{X}$, $\ddot{u}_i$ and $\ddot{U}$ analogously. After eliminating the individual and time effects α_i^* and λ_t from the above regression through the within and time demeaned transformation, we can obtain the two-way within estimator of β^* given by $\hat{\beta}^* = \left(\ddot{X}'\ddot{X}\right)^{-1} \ddot{X}'\ddot{Y}$. Then $\hat{u}_{it}^{(4)}$ can be equivalently represented as

$$\widehat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{x}_{it}^{*\prime}\hat{\beta}^*.$$

Under Assumptions A.1(iv)-(v) and A.2(iii)-(iv), we can readily show that $\hat{\beta}^* - \beta^* = O_P(\eta_{NT})$. Let ϑ_{jl} denotes the (j,l)-th element of $\frac{1}{NT_p}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}} - \mathbf{Z}'\mathbf{Z})$ where $j, l = 1, ..., p$. Then

$$\begin{aligned}
\vartheta_{jl} &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\ddot{u}}_{i,t-j}\widehat{\ddot{u}}_{i,t-l} - \ddot{u}_{i,t-j}\ddot{u}_{i,t-l}\right) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\ddot{u}}_{i,t-j} - \ddot{u}_{i,t-j}\right)\left(\widehat{\ddot{u}}_{i,t-l} - \ddot{u}_{i,t-l}\right) + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\ddot{u}}_{i,t-j} - \ddot{u}_{i,t-j}\right)\ddot{u}_{i,t-l} \\
&\quad + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{u}_{i,t-j}\left(\widehat{\ddot{u}}_{i,t-l} - \ddot{u}_{i,t-l}\right) \\
&\equiv \vartheta_{jl,1} + \vartheta_{jl,2} + \vartheta_{jl,3}.
\end{aligned}$$

Noting that

$$\widehat{\ddot{u}}_{it} = \ddot{y}_{it} - \hat{\beta}^{*\prime}\ddot{x}_{it}^* = \ddot{u}_{it} - (\hat{\beta}^* - \beta^*)'\ddot{x}_{it}^*, \tag{A.22}$$

it is easy to show that $\vartheta_{jl,1} = O_P(\eta_{NT}^2)$. Noting that $\sum_{i=1}^N \ddot{u}_{it} = 0$ for each t, we can apply Assumptions A.1(iii)-(iv) and A.2(iii)-(iv) and show that

$$\begin{aligned}
\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{x}_{i,t-j}^*\ddot{u}_{i,t-l} &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(x_{i,t-j}^* - \bar{x}_{i\cdot}^*\right)\left(u_{i,t-l} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t-l} + \bar{u}_{\cdot\cdot}\right) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(x_{i,t-j}^* - \bar{x}_{i\cdot}^*\right)u_{i,t-l} + O_P\left(\eta_{NT}\right) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}x_{i,t-j}^*u_{i,t-l} + O_P\left(\eta_{NT}\right) = O_P\left(1\right).
\end{aligned}$$

It follows that

$$\vartheta_{jl,2} = -\left(\hat{\beta}^* - \beta^*\right)'\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{x}_{i,t-j}^*\ddot{u}_{i,t-l} = O_P(\eta_{NT})O_P(1) = O_P(\eta_{NT}).$$

Similarly, $\vartheta_{jl,3} = O_P(\eta_{NT})$. Then (i) follows. When $p = 1$, j and l can only take value 1. In this case, $\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{x}_{i,t-j}^*\ddot{u}_{i,t-l} = \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}x_{i,t-1}^*u_{i,t-1} + O_P\left(\eta_{NT}\right) = O_P\left(\eta_{NT}\right)$, $\vartheta_{jl,2} = O_P(\eta_{NT}^2)$, $\vartheta_{jl,3} = O_P(\eta_{NT}^2)$, and $\frac{1}{NT_1}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}} - \mathbf{Z}'\mathbf{Z}) = O_P(\eta_{NT}^2)$.

(ii) The analysis is similar to that in (i) and thus omitted.

(iii) For notational simplicity, we assume that $p = 1$ hereafter. Then we can simply write $\hat{\boldsymbol{\rho}}$, ρ, $\underline{\hat{u}}_{it}^{(l)}$ and $\underline{\ddot{u}}_{i,t}$, as $\hat{\boldsymbol{\rho}}$, ρ, $\hat{u}_{it}^{(l)}$ and $\ddot{u}_{it}$, respectively. Let $\bar{v}_{\cdot t} = \bar{u}_{\cdot t} - \rho\bar{u}_{\cdot,t-1}$. Noting that $\ddot{u}_{it} = u_{it} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t} + \bar{u}_{\cdot\cdot}$ and $\ddot{u}_{i,t-1} = u_{i,t-1} - \bar{u}_{i\cdot} - \bar{u}_{\cdot,t-1} + \bar{u}_{\cdot\cdot}$, we have

$$\begin{aligned}
\ddot{u}_{it} - \rho\ddot{u}_{i,t-1} &= \left(u_{it} - \rho u_{i,t-1}\right) - \left(1-\rho\right)\bar{u}_{i\cdot} - \left(\bar{u}_{\cdot t} - \rho\bar{u}_{\cdot,t-1}\right) + \left(1-\rho\right)\bar{u}_{\cdot\cdot} \\
&= v_{it} - \left(1-\rho\right)\bar{u}_{i\cdot} - \bar{v}_{\cdot t} + \left(1-\rho\right)\bar{u}_{\cdot\cdot}.
\end{aligned}$$

Then

$$\begin{aligned}
\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U}-\rho &= \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}\ddot{u}_{it}-\rho \\
&= \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}\left[v_{it}-(1-\rho)\bar{u}_{i\cdot}-\bar{v}_{\cdot t}+(1-\rho)\bar{u}_{\cdot\cdot}\right] \\
&= \left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}v_{it}-(1-\rho)\left(\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}^{2}\right)^{-1}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}\bar{u}_{i\cdot}
\end{aligned}$$

where the third equality follows from the fact that $\sum_{i=1}^{N}\ddot{u}_{it}=0$ for each t. Noting that $\sum_{t=1}^{T}\ddot{u}_{it}=0$ for each i, we have by Assumptions A.1(i), (iv) and A.2(iii)-(iv)

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}\bar{u}_{i\cdot} &= \frac{1}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\ddot{u}_{it}-\ddot{u}_{iT}\right)\bar{u}_{i\cdot}=-\frac{1}{NT}\sum_{i=1}^{N}\ddot{u}_{iT}\bar{u}_{i\cdot} \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\left(u_{iT}-\bar{u}_{i\cdot}-\bar{u}_{\cdot T}+\bar{u}_{\cdot\cdot}\right)\bar{u}_{i\cdot} \\
&= O_P\left(T^{-1}\right)+O_P\left(T^{-2}\right)+O_P\left(N^{-1}T^{-1}\right)+O_P\left(N^{-1}T^{-2}\right)=O_P\left(T^{-1}\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\ddot{u}_{i,t-1}v_{it} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}u_{i,t-1}v_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\bar{u}_{i\cdot}v_{it}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\bar{u}_{\cdot,t-1}v_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\bar{u}_{\cdot\cdot}v_{it} \\
&= O_P\left((NT)^{-1/2}\right)+O_P\left(T^{-1}\right)+O_P\left(N^{-1}\right)+O_P\left(N^{-1}T^{-1}\right) \\
&= O_P\left((NT)^{-1/2}+N^{-1}+T^{-1}\right).
\end{aligned}$$

In sum, we have $\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U}-\rho=O_P(\eta_{NT})$. ■

To simply the proof, we assume that $p=1$ hereafter.

Lemma A.8 *Let $\breve{x}_{it}^{(l)}=x_{it}^{(l)}-\rho x_{i,t-1}^{(l)}$ and $K_{lNT}=\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\breve{x}_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U$ for $l=1,2,3,4$. Suppose that Assumptions A.1(iv)-(v), A.2(iii)-(iv), A.4(iv), and A.5(iii)-(iv) hold. Then*

(i) $K_{1NT}=O_P\left((NT)^{-1}\right)$,

(ii) $K_{2NT}=\frac{1-\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\bar{u}_{i\cdot}+O_P((NT)^{-1}+T^{-2})$,

(iii) $K_{3NT}=\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}[(1-\rho L)\bar{u}_{\cdot t}]+O_P((NT)^{-1}+N^{-2})$,

(iv) $K_{4NT}=\frac{1-\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\bar{u}_{i\cdot}+\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}[(1-\rho L)\bar{u}_{\cdot t}]+O_P(N^{-2}+T^{-2})$, $K_{4NT}-K_{2NT}=\frac{1-\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}[(1-\rho L)\bar{u}_{\cdot t}]+O_P((NT)^{-1}+N^{-2})$, and $K_{4NT}-K_{3NT}=\frac{1-\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\bar{u}_{i\cdot}+O_P((NT)^{-1}+T^{-2})$.

Proof. (i) Noting that $\breve{x}_{it}^{(1)}=x_{it}-\rho x_{i,t-1}\equiv\breve{x}_{it}$, we can readily apply Assumptions A.1(iv)-(v) and A.4(v) to show that

$$\begin{aligned}
K_{1NT} &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\breve{x}_{it}'\left(X'X\right)^{-1}X'U \\
&= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\left(x_{it}-\rho x_{i,t-1}\right)'\left(\frac{1}{NT}X'X\right)^{-1}\frac{1}{NT}X'U \\
&= O_P\left((NT)^{-1/2}\right)O_P(1)\,O_P\left((NT)^{-1/2}\right)=O_P\left((NT)^{-1}\right).
\end{aligned}$$

(ii) Note that $\breve{x}_{it}^{(2)} = x_{it}^{(2)} - \rho x_{i,t-1}^{(2)} = \left((x_{it} - \rho x_{i,t-1})', (d_{\alpha,it} - \rho d_{\alpha,it-1})' \right)' \equiv \left(\breve{x}_{it}', \breve{d}_{\alpha,it}' \right)'$. By Lemma A.1 with $D = D_\alpha$, we have

$$
\begin{aligned}
K_{2NT} &= \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{x}_{it}^{(2)\prime} \left(X^{(2)\prime} X^{(2)} \right)^{-1} X^{(2)\prime} U \\
&= \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \left(\breve{x}_{it}', \breve{d}_{\alpha,it}' \right) \begin{pmatrix} X_{D_\alpha}^* & -X_{D_\alpha}^* B_\alpha \\ -B_\alpha' X_{D_\alpha}^* & (D_\alpha' D_\alpha)^{-1} + B_\alpha' X_D^* B_\alpha \end{pmatrix} \begin{pmatrix} X'U \\ D_\alpha' U \end{pmatrix} \\
&= \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \Big\{ \breve{x}_{it}' X_{D_\alpha}^* X'U - \breve{d}_{\alpha,it}' B_\alpha' X_{D_\alpha}^* X'U - \breve{x}_{it}' X_{D_\alpha}^* B_\alpha D_\alpha' U + \breve{d}_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} D_\alpha' U \\
&\quad + \breve{d}_{\alpha,it}' B_\alpha' X_{D_\alpha}^* B_\alpha D_\alpha' U \Big\} \\
&\equiv K_{2NT,1} - K_{2NT,2} - K_{2NT,3} + K_{2NT,4} + K_{2NT,5}, \text{ say.}
\end{aligned}
$$

As in (i), we can show that $K_{2NT,1} = O_P((NT)^{-1})$ by Assumption A.1(iv)-(v). Noting that $d_{\alpha,it}' \iota_{N-1} = 1$ for $i \le N-1$ and $d_{\alpha,Nt} = -\iota_{N-1}$, we have by (A.1)

$$
\begin{aligned}
& \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} D_\alpha' X \\
= & \frac{1}{NTT_1} \sum_{i=1}^{N} v_{it} d_{\alpha,it}' \left[\left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}', -\frac{1}{N} \iota_{N-1} \right) \otimes \iota_T' \right] X \\
= & \frac{1}{NTT_1} \sum_{i=1}^{N-1} \sum_{t=2}^{T} v_{it} \left[\left(d_{\alpha,it}' - \frac{1}{N} \iota_{N-1}', -\frac{1}{N} \right) \otimes \iota_T' \right] X + \frac{1}{NTT_1} \sum_{t=2}^{T} v_{Nt} \left(\left(-\frac{1}{N} \iota_{N-1}', \frac{N-1}{N} \right) \otimes \iota_T' \right) X \\
= & \frac{1}{NT_1} \sum_{i=1}^{N-1} \sum_{t=2}^{T} v_{it} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})' + \frac{1}{NT_1} \sum_{t=2}^{T} v_{N1} (\bar{x}_{N\cdot} - \bar{x}_{\cdot\cdot})' \\
= & \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})'
\end{aligned}
$$

and similarly

$$
\begin{aligned}
& \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d_{\alpha,it-1}' (D_\alpha' D_\alpha)^{-1} D_\alpha' X \\
= & \frac{1}{NTT_1} \sum_{i=1}^{N} v_{it} d_{\alpha,it-1}' \left[\left(I_{N-1} - \frac{1}{N} \iota_{N-1} \iota_{N-1}', -\frac{1}{N} \iota_{N-1} \right) \otimes \iota_T' \right] X \\
= & \frac{1}{NTT_1} \sum_{i=1}^{N-1} \sum_{t=2}^{T} v_{it} \left[\left(d_{\alpha,it-1}' - \frac{1}{N} \iota_{N-1}', -\frac{1}{N} \right) \otimes \iota_T' \right] X + \frac{1}{NTT_1} \sum_{t=2}^{T} v_{Nt} \left(\left(-\frac{1}{N} \iota_{N-1}', \frac{N-1}{N} \right) \otimes \iota_T' \right) X \\
= & \frac{1}{NT_1} \sum_{i=1}^{N-1} \sum_{t=2}^{T} v_{it} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})' + \frac{1}{NT_1} \sum_{t=2}^{T} v_{N1} (\bar{x}_{N\cdot} - \bar{x}_{\cdot\cdot})' \\
= & \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})'.
\end{aligned}
$$

Then $\frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{d}_{\alpha,it}' B_\alpha' = \frac{1-\rho}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})' = O_P(T^{-1} + (NT)^{-1/2})$ by Assumptions

A.4(iv) and A.5(iii) and

$$\begin{aligned} K_{2NT,2} &\leq \left\| \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{d}'_{\alpha,it} B'_{\alpha} \right\| \left\| \left(\frac{1}{NT} X' M_{D_\alpha} X \right)^{-1} \right\| \left\| \frac{1}{NT} X'U \right\| \\ &= O_P(T^{-1} + (NT)^{-1/2}) O_P(1) O_P((NT)^{-1/2}) = O_P\left(N^{-1/2} T^{-3/2} + (NT)^{-1} \right). \end{aligned}$$

Similarly, we can show that

$$\begin{aligned} K_{2NT,3} &\leq \left\| \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{x}'_{it} \right\| \left\| \left(\frac{1}{NT} X' M_{D_\alpha} X \right)^{-1} \right\| \left\| \frac{1}{NT} B_\alpha D'_\alpha U \right\| \\ &= O_P\left((NT)^{-1/2} \right) O_P(1) O_P(T^{-1} + (NT)^{-1/2}) = O_P(N^{-1/2} T^{-3/2} + (NT)^{-1}), \\ K_{2NT,5} &\leq \left\| \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{d}'_{\alpha,it} B'_{\alpha} \right\| \left\| \left(\frac{1}{NT} X' M_{D_\alpha} X \right)^{-1} \right\| \left\| \frac{1}{NT} B_\alpha D'_\alpha U \right\| \\ &= O_P\left(T^{-1} + (NT)^{-1/2} \right) O_P(1) O_P(T^{-1} + (NT)^{-1/2}) = O_P(T^{-2} + N^{-1} T^{-1}). \end{aligned}$$

and

$$\begin{aligned} K_{2NT,4} &= \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \breve{d}'_{\alpha,it} \left(D'_\alpha D_\alpha \right)^{-1} D'_\alpha U = \frac{1-\rho}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \left(\bar{u}_{i\cdot} - \bar{u}_{\cdot\cdot} \right) \\ &= \frac{1-\rho}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \bar{u}_{i\cdot} + O_P((NT)^{-1}). \end{aligned}$$

It follows that $K_{2NT} = \frac{1-\rho}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \bar{u}_{i\cdot} + O_P((NT)^{-1} + T^{-2})$ as $(NT)^{-1} + T^{-2} \geq 2N^{-1/2} T^{-3/2}$ by the Cauchy-Schwarz inequality.

(iii) The proof is analogous to that of (ii). The major difference is that we need to use the fact that $d'_{\lambda,it} \iota_{T-1} = 1$ for $t \leq T-1$ and $d_{\alpha,iT} = -\iota_{T-1}$, and that by (A.6)

$$\begin{aligned} &\frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d'_{\lambda,it} \left(D'_\lambda D_\lambda \right)^{-1} D'_\lambda X \\ =\;& \frac{1}{N^2 T_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d'_{\lambda,it} \left[\iota'_N \otimes \left(I_{T-1} - \frac{1}{T} \iota_{T-1} \iota'_{T-1}, -\frac{1}{T} \iota_{T-1} \right) \right] X \\ =\;& \frac{1}{N^2 T_1} \sum_{i=1}^{N} \sum_{t=2}^{T-1} v_{it} \left[\iota'_N \otimes \left(d'_{\lambda,it} - \frac{1}{T} \iota'_{T-1}, -\frac{1}{T} \right) \right] X + \frac{1}{N^2 T_1} \sum_{i=1}^{N} v_{iT} \left[\iota'_N \otimes \left(-\frac{1}{T} \iota'_{T-1}, \frac{T-1}{T} \right) \right] X \\ =\;& \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T-1} v_{it} \left(\bar{x}_{\cdot t} - \bar{x}_{\cdot\cdot} \right)' + \frac{1}{NT_1} \sum_{i=1}^{N} v_{iT} \left(\bar{x}_{\cdot T} - \bar{x}_{\cdot\cdot} \right)' = \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \left(\bar{x}_{\cdot t} - \bar{x}_{\cdot\cdot} \right)', \end{aligned}$$

and

$$\begin{aligned} \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d'_{\lambda,it-1} \left(D'_\lambda D_\lambda \right)^{-1} D'_\lambda X &= \frac{1}{N^2 T_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} d'_{\lambda,it-1} \left[\iota'_N \otimes \left(I_{T-1} - \frac{1}{T} \iota_{T-1} \iota'_{T-1}, -\frac{1}{T} \iota_{T-1} \right) \right] X \\ &= \frac{1}{N^2 T_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \left[\iota'_N \otimes \left(d'_{\lambda,it-1} - \frac{1}{T} \iota'_{T-1}, -\frac{1}{T} \right) \right] X \\ &= \frac{1}{NT_1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{it} \left(\bar{x}_{\cdot t-1} - \bar{x}_{\cdot\cdot} \right)'. \end{aligned}$$

The dominant term then becomes

$$\begin{aligned}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\breve{d}_{\lambda,it}'\left(D_\lambda' D_\lambda\right)^{-1}D_\lambda' U &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\left[\left(\bar{u}_{\cdot t}-\bar{u}_{\cdot\cdot}\right)-\rho\left(\bar{u}_{\cdot t-1}-\bar{u}_{\cdot\cdot}\right)\right]\\ &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\left(1-\rho L\right)\bar{u}_{\cdot t}+O_P((NT)^{-1}).\end{aligned}$$

(iv) The proof is a combination of (ii)-(iii) as in that of Lemma A.5. ∎

Lemma A.9 *Let* $\breve{x}_{it}^{(l)}=x_{it}^{(l)}-\rho x_{i,t-1}^{(l)}$ *and* $L_{lNT}=\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\breve{x}_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right)^2$ *for* $l=1,2,3,4$. *Suppose that Assumptions A.1(iv)-(v), A.2(iii)-(iv), A.4(iv), and A.5(iii)-(iv) hold. Then*

(i) $L_{1NT}=O_P\left((NT)^{-1}\right)$,

(ii) $L_{2NT}=(1-\rho)^2\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+O_P\left((NT)^{-1}+T^{-2}\right)$,

(iii) $L_{3NT}=\frac{1}{T}\sum_{t=1}^{T}[(1-\rho L)\bar{u}_{\cdot t}]^2+O_P((NT)^{-1}+N^{-2})$,

(iv) $L_{4NT}=(1-\rho)^2\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+\frac{1}{T}\sum_{t=1}^{T}[(1-\rho L)\bar{u}_{\cdot t}]^2+O_P\left(T^{-2}+N^{-2}\right)$, $L_{4NT}-L_{2NT}=\frac{1}{T}\sum_{t=1}^{T}[(1-\rho L)\bar{u}_{\cdot t}]^2+O_P\left((NT)^{-1}+N^{-2}\right)$, *and* $L_{4NT}-L_{3NT}=(1-\rho)^2\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+O_P\left((NT)^{-1}+T^{-2}\right)$,

Proof. (i) Noting that $\breve{x}_{it}^{(1)}=x_{it}-\rho x_{i,t-1}\equiv\breve{x}_{it}$, we can readily apply Assumption A.1(iv)-(v) to show that

$$\begin{aligned}L_{1NT} &= U'X\left(X'X\right)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}\breve{x}_{it}'\left(X'X\right)^{-1}X'U\\ &\leq \left\|\left(\frac{1}{NT}X'X\right)^{-1}\right\|^2\left\|\frac{1}{NT}X'U\right\|^2\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}\breve{x}_{it}'\right\|=O_P\left((NT)^{-1}\right).\end{aligned}$$

(ii) By Lemma A.1 with $D=D_\alpha$ and using $\breve{x}_{it}^{(2)}\breve{x}_{it}^{(2)\prime}=\begin{pmatrix}\breve{x}_{it}\breve{x}_{it}' & \breve{x}_{it}\breve{d}_{\alpha,it}'\\ \breve{d}_{\alpha,it}\breve{x}_{it}' & \breve{d}_{\alpha,it}\breve{d}_{\alpha,it}'\end{pmatrix}$, we have

$$\begin{aligned}L_{2NT} &= U'X^{(2)}\left(X^{(2)\prime}X^{(2)}\right)^{-1}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}^{(2)}\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\\ &= \left(U'X,U'D_\alpha\right)\begin{pmatrix}X_{D_\alpha}^{*} & -X_{D_\alpha}^{*}B_\alpha\\ -B_\alpha'X_{D_\alpha}^{*} & \left(D_\alpha'D_\alpha\right)^{-1}+B_\alpha'X_D^{*}B_\alpha\end{pmatrix}\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}^{(2)}\breve{x}_{it}^{(2)\prime}\\ &\quad\times\begin{pmatrix}X_{D_\alpha}^{*} & -X_{D_\alpha}^{*}B_\alpha\\ -B_\alpha'X_{D_\alpha}^{*} & \left(D_\alpha'D_\alpha\right)^{-1}+B_\alpha'X_D^{*}B_\alpha\end{pmatrix}\begin{pmatrix}X'U\\ D_\alpha'U\end{pmatrix}\\ &= \left(\zeta_1',\zeta_2'\right)\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}^{(2)}\breve{x}_{it}^{(2)\prime}\begin{pmatrix}\zeta_1\\ \zeta_2\end{pmatrix}\\ &= \zeta_1'\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}\breve{x}_{it}'\zeta_1+\zeta_2'\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{d}_{\alpha,it}\breve{d}_{\alpha,it}'\zeta_2+2\zeta_1'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\breve{x}_{it}\breve{d}_{\alpha,it}'\zeta_2\\ &\equiv L_{2NT,1}+L_{2NT,2}+2L_{2NT,3},\ \text{say},\end{aligned}$$

where $\zeta_1=X_{D_\alpha}^{*\prime}X'U-X_{D_\alpha}^{*\prime}B_\alpha D_\alpha'U$ and $\zeta_2=-B_\alpha'X_{D_\alpha}^{*\prime}X'U+\left(D_\alpha'D_\alpha\right)^{-1}D_\alpha'U+B_\alpha'X_{D_\alpha}^{*\prime}B_\alpha D_\alpha'U$. It is easy to show that $L_{2NT,1} = O_P((NT)^{-1} + T^{-2})$ by Assumptions A.1(iv)-(v) and Lemma A.4(i). For $L_{2NT,2}$,

$$\begin{aligned} L_{2NT,2} &= U'XX^*_{D_\alpha}B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha X^{*\prime}_{D_\alpha}X'U \\ &\quad +U'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1}\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha U \\ &\quad +U'D_\alpha B'_\alpha X^*_{D_\alpha}B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha X^{*\prime}_{D_\alpha}B_\alpha D'_\alpha U \\ &\quad -2U'XX^*_{D_\alpha}B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha U \\ &\quad -2U'XX^*_{D_\alpha}B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha X^{*\prime}_{D_\alpha}B_\alpha D'_\alpha U \\ &\quad +2U'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1}\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha X^{*\prime}_{D_\alpha}B_\alpha D'_\alpha U \\ &\equiv L_{2NT,21}+L_{2NT,22}+L_{2NT,23}-2L_{2NT,24}-2L_{2NT,25}+2L_{2NT,26}, \text{ say.} \end{aligned}$$

Noting that $\breve{d}'_{\alpha,it}B'_\alpha = \breve{d}'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha X = (1-\rho)\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)'$, we have

$$B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha = (1-\rho)^2\frac{1}{N}\sum_{i=1}^N\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)' = O_P(1).$$

This, in conjunction with Assumption A.1(iv)-(v) and Lemma A.5(i), implies that

$$\begin{aligned} L_{2NT,21} &\leq \left\|B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha\right\|\left\|\left(\frac{1}{NT}X'M_{D_\alpha}X\right)^{-1}\right\|^2\left\|\frac{1}{NT}X'U\right\|^2 = O_P((NT)^{-1}), \\ L_{2NT,23} &\leq \left\|B_\alpha \frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}B'_\alpha\right\|\left\|\left(\frac{1}{NT}X'M_{D_\alpha}X\right)^{-1}\right\|^2\left\|\frac{1}{NT}B_\alpha D'_\alpha U\right\|^2 = O_P(T^{-2}+(NT)^{-1}). \end{aligned}$$

Noting that $\breve{d}'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha U = (1-\rho)\left(\bar{u}_{i\cdot}-\bar{u}_{\cdot\cdot}\right)$, we have

$$\begin{aligned} L_{2NT,22} &= U'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1}\frac{1}{N}\sum_{i=1}^N\sum_{t=2}^T \breve{d}_{\alpha,it}\breve{d}'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha U = (1-\rho)^2\frac{1}{N}\sum_{i=1}^N\left(\bar{u}_{i\cdot}-\bar{u}_{\cdot\cdot}\right)^2 \\ &= (1-\rho)^2\frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P((NT)^{-1}). \end{aligned}$$

Analogously, we can show that $L_{2NT,2j} = O_P(T^{-2}+(NT)^{-1})$ for $j=4,5,6$ and $L_{2NT,3} = O_P(T^{-2}+(NT)^{-1})$. It follows that $L_{2NT} = (1-\rho)^2\frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P\left((NT)^{-1}+T^{-2}\right)$.

(iii) The proof is analogous to that of (ii) with the major difference as outlined in the proof of Lemma A.8(iii).

(iv) The proof is a combination of (ii) and (iii) as in that of Lemma A.5(iv) and thus omitted. ■

Proof of Theorem 2.2. Again, we assume that $p=1$. Noting that

$$\begin{aligned} \hat{\rho}-\rho &= \left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\rho \\ &= \left[\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\hat{\mathbf{Z}}'\hat{\mathbf{U}}+\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right)+\left[\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U}-\rho\right], \end{aligned}$$

we have $\hat{\rho}-\rho=O_P(\eta_{NT})$ by Lemma A.7 and the triangle inequality. Noting that $(y_{it}-\hat{\rho}y_{i,t-1})-(\hat{y}_{it}^{(m)}-\hat{\rho}'\hat{y}_{i,t-1}^{(m)})=(y_{it}-\rho y_{i,t-1})-(\hat{y}_{it}^{(m)}-\rho\hat{y}_{i,t-1}^{(m)})+(\hat{\rho}-\rho)(\hat{y}_{i,t-1}^{(m)}-y_{i,t-1})$, we have

$$
\begin{aligned}
CV^*(m) &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(y_{it}-\hat{\rho}y_{i,t-1})-(\hat{y}_{it}^{(m)}-\hat{\rho}'\hat{y}_{i,t-1}^{(m)})\right]^2 \\
&= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(y_{it}-\rho y_{i,t-1})-(\hat{y}_{it}^{(m)}-\rho\hat{y}_{i,t-1}^{(m)})\right]^2+\frac{(\hat{\rho}-\rho)^2}{NT_p}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{y}_{i,t-1}^{(m)}-y_{i,t-1})^2 \\
&\quad +\frac{2(\hat{\rho}-\rho)}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{y}_{i,t-1}^{(m)}-y_{i,t-1})\left[(y_{it}-\rho y_{i,t-1})-(\hat{y}_{it}^{(m)}-\rho\hat{y}_{i,t-1}^{(m)})\right] \\
&= CV_1^*(m)+CV_2^*(m)+CV_3^*(m).
\end{aligned}
$$

As in the proof of Theorem 2.1, we will use $CV_{l,m}^*$ and $CV_{l,m}^*(j)$ to denote $CV^*(m)$ and $CV_j^*(m)$ when the true model is Model l. Note that $CV_{l,m}^*=\sum_{j=1}^{3}CV_{l,m}^*(j)$.

Case 1: Model 1 is the true model. In this case, Models 2-4 are all overfitted models and we will show that $P\left(CV_{1,1}^*<CV_{1,m}^*\right)\to 1$ for $m=2,3,4$. When Model 1 is the true model, we have by (A.11)

$$
\begin{aligned}
&(y_{it}-\hat{y}_{it}^{(l)})-\rho(y_{i,t-1}-\hat{y}_{i,t-1}^{(l)}) \\
&= \frac{1}{1-h_{it}^{(l)}}\left[u_{it}-x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]-\frac{\rho}{1-h_{i,t-1}^{(l)}}\left[u_{i,t-1}-x_{i,t-1}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right] \\
&= c_{it,l}\left[v_{it}-\breve{x}_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]+\rho\varkappa_{it,l}\left[u_{i,t-1}-x_{i,t-1}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]. \qquad (A.23)
\end{aligned}
$$

where $\breve{x}_{it}^{(l)}=x_{it}^{(l)}-\rho x_{i,t-1}^{(l)}$, $c_{it,l}=(1-h_{it}^{(l)})^{-1}$, and $\varkappa_{it,l}=c_{it,l}-c_{it-1,l}$ for $l=1,2,3,4$. By Lemma A.6, we have

$$\max_{i,t}\varkappa_{it,l}=o_P(\delta_{NT}) \text{ for } l=1,2,3,4. \qquad (A.24)$$

Note that

$$
\begin{aligned}
CV_{1,l}^*(1) &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}c_{it,l}^2\left[v_{it}-\breve{x}_{it}^{(l)\prime}(X^{(l)\prime}X^{(l)})^{-1}X^{(l)\prime}U\right]^2 \\
&\quad +\frac{\rho^2}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\varkappa_{it,l}^2\left[u_{i,t-1}-x_{i,t-1}^{(l)\prime}(X^{(l)\prime}X^{(l)})^{-1}X^{(l)\prime}U\right]^2 \\
&\quad +\frac{2\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}c_{it,l}\varkappa_{it,l}\left[v_{it}-\breve{x}_{it}^{(l)\prime}(X^{(l)\prime}X^{(l)})^{-1}X^{(l)\prime}U\right]\left[u_{i,t-1}-x_{i,t-1}^{(l)\prime}(X^{(l)\prime}X^{(l)})^{-1}X^{(l)\prime}U\right] \\
&\equiv CV_{1,l}^*(1,1)+\rho^2CV_{1,l}^*(1,2)-2\rho CV_{1,l}^*(1,3),\ \text{say.}
\end{aligned}
$$

We first study $CV_{1,2}^*(1)-CV_{1,1}^*(1)$. Following the study of $CV_{1,2}-CV_{1,1}$ in the proof of Theorem 2.1, we can readily apply Lemmas A.8(i)-(ii) and A.9(i)-(ii), Assumptions A.4(ii) and A.5(i) to show that

$$\begin{aligned}
T_1\left[CV_{1,2}^*(1,1)-CV_{1,1}^*(1,1)\right] &= \frac{2}{NT_1}\sum_{i=1}^N\sum_{t=2}^T v_{it}^2+\frac{T_1(1-\rho)^2}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2-\frac{2(1-\rho)}{N}\sum_{i=1}^N\sum_{t=2}^T v_{it}\bar{u}_{i\cdot}+o_P(1)\\
&= \frac{2}{NT_1}\sum_{i=1}^N\sum_{t=2}^T v_{it}^2+\frac{T_1(1-\rho)^2}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2-\frac{2T_1(1-\rho)}{N}\sum_{i=1}^N \bar{v}_{i\cdot}\bar{u}_{i\cdot}+o_P(1)\\
&= \left(\frac{2}{NT_1}\sum_{i=1}^N\sum_{t=2}^T v_{it}^2-\frac{T_1}{N}\sum_{i=1}^N \bar{v}_{i\cdot}^2\right)+\frac{T_1}{N}\sum_{i=1}^N\left[\bar{v}_{i\cdot}-(1-\rho)\bar{u}_{i\cdot}\right]^2+o_P(1)\\
&\xrightarrow{P} 2\bar{\sigma}_v^2-\bar{\sigma}_{v1}^2,
\end{aligned}$$

where we use the fact that $\bar{v}_{i\cdot}=\frac{1}{T_1}\sum_{t=2}^T v_{it}=\frac{1}{T_1}\sum_{t=2}^T(u_{it}-\rho u_{i,t-1})=(1-\rho)\bar{u}_{i\cdot}+O_P\left(T^{-1}\right)$. Similarly, using (A.24) and following the analysis of $CV_{1,2}-CV_{1,1}$, we can readily show that $T_1[CV_{1,2}^*(1,2)-CV_{1,1}^*(1,2)]=o_P(1)$ and $T_1\left[CV_{1,2}^*(1,3)-CV_{1,1}^*(1,3)\right]=o_P(1)$. It follows that $T_1\left[CV_{1,2}^*(1)-CV_{1,1}^*(1)\right]\xrightarrow{P} 2\bar{\sigma}_v^2-\bar{\sigma}_{v1}^2$.

By (A.11) and (A.23),

$$\begin{aligned}
CV_{1,l}^*(2) &= (\hat{\rho}-\rho)^2\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T\left(\hat{y}_{i,t-1}^{(l)}-y_{i,t-1}\right)^2\\
&= (\hat{\rho}-\rho)^2\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T c_{it,l}^2\left[u_{it}-x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]^2\equiv(\hat{\rho}-\rho)^2 D_{1,l}(1),\text{ and}\\
CV_{1,l}^*(3) &= \frac{1}{NT_1}(\hat{\rho}-\rho)\sum_{i=1}^N\sum_{t=2}^T\left(\hat{y}_{i,t-1}^{(l)}-y_{i,t-1}\right)\left[y_{it}-\hat{y}_{it}^{(l)}-\rho(y_{i,t-1}-\hat{y}_{i,t-1}^{(l)})\right]\\
&= (\hat{\rho}-\rho)\left\{\frac{1}{NT_1}\sum_{i=1}^N\sum_{t=2}^T c_{it,l}^2\left[u_{it}-x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]\left[v_{it}-\breve{x}_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]\right.\\
&\quad\left.+\frac{\rho}{NT_1}\sum_{i=1}^N\sum_{t=2}^T c_{it,l}\varkappa_{it,l}\left[u_{it}-x_{it}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]\left[u_{i,t-1}-x_{i,t-1}^{(l)\prime}\left(X^{(l)\prime}X^{(l)}\right)^{-1}X^{(l)\prime}U\right]\right\}\\
&\equiv (\hat{\rho}-\rho)\left\{D_{1,l}(2)+D_{1,l}(3)\right\},\text{ say.}
\end{aligned}$$

As in the analysis of $CV_{1,2}-CV_{1,1}$, we cam readily show that $D_{1,2}(1)-D_{1,1}(1)=O_P\left(T^{-1}\right)$, $D_{1,l}(\ell)=O_P\left((NT)^{-1}\right)$ and $D_{2,l}(\ell)=O_P\left(T^{-1}\right)$ for $\ell=2,3$. Then

$T_1\left[CV_{1,2}^*(3)-CV_{1,1}^*(3)\right]=(\hat{\rho}-\rho)^2 O_P(1)=o_P(1)$ and $T_1\left[CV_{1,2}^*(3)-CV_{1,1}^*(3)\right]=(\hat{\rho}-\rho)O_P(1)=o_P(1)$.

In sum, we have

$$\begin{aligned}
T_1\left[CV_{1,2}^*-CV_{1,1}^*\right] &= \left(\frac{2}{NT_1}\sum_{i=1}^N\sum_{t=2}^T v_{it}^2-\frac{T_1}{N}\sum_{i=1}^N \bar{v}_{i\cdot}^2\right)+\frac{T_1}{N}\sum_{i=1}^N\left[\bar{v}_{i\cdot}-(1-\rho)\bar{u}_{i\cdot}\right]^2+o_P(1)\\
&\xrightarrow{P} 2\bar{\sigma}_v^2-\bar{\sigma}_{v1}^2. \qquad (A.25)
\end{aligned}$$

Similarly, by using Lemma A.8(i) and (iii), Lemma A.9(i) and (iii), Assumptions A.4(ii) and A.5(ii) we can show that

$$\begin{aligned}
T_1\left[CV_{1,3}^*-CV_{1,1}^*\right] &= \left(\frac{2}{NT_1}\sum_{i=1}^N\sum_{t=2}^T v_{it}^2-\frac{N}{T_1}\sum_{t=2}^T \bar{v}_{\cdot t}^2\right)+\frac{N}{T_1}\sum_{t=2}^T\left[\bar{v}_{\cdot t}-(1-\rho L)\bar{u}_{\cdot t}\right]^2+o_P(1)\\
&\xrightarrow{P} 2\bar{\sigma}_v^2-\bar{\sigma}_{v2}^2, \qquad (A.26)
\end{aligned}$$

where we use the fact that $\bar{v}_{\cdot t} = \frac{1}{N}\sum_{i=1}^{N} v_{it} = \frac{1}{N}\sum_{i=1}^{N} \Phi(L) u_{it} = \Phi(L)\bar{u}_{\cdot t}$. By using Lemma A.8(iv) and Lemma A.9(i) and (iv),

$$\begin{aligned}(N \wedge T_1)[CV_{1,4} - CV_{1,1}] &= (N \wedge T_1)\left\{\left(T_1^{-1} + N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T} v_{it}^2 - \frac{1}{N}\sum_{i=1}^{N}\bar{v}_{i\cdot}^2 - \frac{1}{T_1}\sum_{t=2}^{T}\bar{v}_{\cdot t}^2\right\} + o_P(1) \\ &\xrightarrow{P} 2(1+c)\bar{\sigma}_v^2 - \left(\bar{\sigma}_{v1}^2 + c\bar{\sigma}_{v2}^2\right)1\{c_1 \geq 1\} - \left(c\bar{\sigma}_{v1}^2 + \bar{\sigma}_{v2}^2\right)1\{c_1 < 1\}, \quad \text{(A.27)}\end{aligned}$$

where $c = \lim_{(N,T)\to\infty}\left(\frac{N}{T} \wedge \frac{T}{N}\right)$ and $c_1 = \lim_{(N,T)\to\infty}\frac{N}{T}$. Combining (A.25)-(A.27) yields $P\left(CV_{1,1}^* < CV_{1,m}^*\right) \to 1$ for $m = 2,3,4$ provided $\max\left(\bar{\sigma}_{v1}^2, \bar{\sigma}_{u2}^2\right) < 2\bar{\sigma}_v^2$.

Case 2: Model 2 is the true model. In this case, Models 1 and 2 are underfitted and Model 4 is overfitted and we will show that $P\left(CV_{2,2}^* < CV_{2,m}^*\right) \to 1$ for $m = 1,3,4$. Let $u_{\alpha,it}$ and U_α be as defined in the proof of Theorem 2.1. Following the steps to obtain (A.15), we can show that

$$\begin{aligned}&(y_{it} - \hat{y}_{it}^{(1)}) - \rho(y_{i,t-1} - \hat{y}_{i,t-1}^{(1)}) \\ =\ & \frac{1}{1 - h_{it}^{(1)}}\left[u_{\alpha,it} - x_{it}'(X'X)^{-1}X'U_\alpha\right] - \frac{\rho}{1 - h_{i,t-1}^{(1)}}\left[u_{\alpha,it-1} - x_{i,t-1}'(X'X)^{-1}X'U_\alpha\right] \\ =\ & c_{it,1}\left[(1-\rho)\alpha_i + v_{it} - \breve{x}_{it}'(X'X)^{-1}X'U_\alpha\right] + \rho\varkappa_{it,1}\left[u_{\alpha,it-1} - x_{i,t-1}'(X'X)^{-1}X'U_\alpha\right]. \quad \text{(A.28)}\end{aligned}$$

Then

$$\begin{aligned}CV_{2,1}^* &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T} c_{it,1}^2\left[(1-\rho)\alpha_i + v_{it} - \breve{x}_{it}'(X'X)^{-1}X'U_\alpha\right]^2 \\ &\quad + \frac{\rho^2}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\varkappa_{it,1}^2\left[u_{\alpha,it-1} - x_{i,t-1}'(X'X)^{-1}X'U_\alpha\right]^2 \\ &\quad + \frac{2\rho}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T} c_{it,1}\varkappa_{it,1}\left[(1-\rho)\alpha_i + v_{it} - \breve{x}_{it}'(X'X)^{-1}X'U_\alpha\right]\left[u_{\alpha,it-1} - x_{i,t-1}'(X'X)^{-1}X'U_\alpha\right] \\ &\equiv D_{2,1}(1) + \rho^2 D_{2,1}(2) + 2\rho D_{2,1}(3), \text{ say.}\end{aligned}$$

It is easy to show that by Assumptions A.1(i) and (iv)-(vi), A.4(ii), and A.6(i)

$$D_{2,1}(1) = \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(1-\rho)\alpha_i - \breve{x}_{it}'(X'X)^{-1}X'D_\alpha\underline{\alpha}\right]^2 + \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T} v_{it}^2 + o_P(1) \xrightarrow{P} c_{\alpha,X}^* + \bar{\sigma}_v^2.$$

In addition, $D_{2,\ell}(2) = o_P(1)$ for $\ell = 2,3$. Thus $CV_{2,1}^* = c_{\alpha,X}^* + \bar{\sigma}_v^2$. Following the analysis in Case 1 and noting that

$$\begin{aligned}&(y_{it} - \hat{y}_{it}^{(2)}) - \rho(y_{i,t-1} - \hat{y}_{i,t-1}^{(2)}) \\ =\ & \frac{1}{1 - h_{it}^{(2)}}\left[u_{it} - x_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right] - \frac{\rho}{1 - h_{i,t-1}^{(2)}}\left[u_{it-1} - x_{i,t-1}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right] \\ =\ & c_{it,2}\left[v_{it} - \breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right] + \rho\varkappa_{it,2}\left[u_{it-1} - x_{i,t-1}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right],\end{aligned}$$

we can readily show that $CV_{2,2}^* = \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T} v_{it}^2 + o_P(1) \xrightarrow{P} \bar{\sigma}_v^2$. It follows that

$$CV_{2,1}^* - CV_{2,2}^* \xrightarrow{P} c_{\alpha,X}^* > 0. \quad \text{(A.29)}$$

To study $CV^*_{2,3}$, noting that

$$\begin{aligned}
& (y_{it}-\hat{y}^{(3)}_{it})-\rho(y_{i,t-1}-\hat{y}^{(3)}_{i,t-1}) \\
= & \frac{1}{1-h^{(3)}_{it}}[u_{\alpha,it}-x^{(3)\prime}_{it}(X^{(3)\prime}X^{(3)})^{-1}X^{(3)\prime}U_\alpha]-\frac{\rho}{1-h^{(3)}_{i,t-1}}[u_{\alpha,it-1}-x^{(3)\prime}_{i,t-1}(X^{(3)\prime}X^{(3)})^{-1}X^{(3)\prime}U_\alpha] \\
= & c_{it,3}[(1-\rho)\alpha_i+v_{it}-\breve{x}^{(3)\prime}_{it}(X^{(3)\prime}X^{(3)})^{-1}X^{(3)\prime}U_\alpha]+\rho\varkappa_{it,3}[u_{\alpha,it-1}-x^{(3)\prime}_{i,t-1}(X^{(3)\prime}X^{(3)})^{-1}X^{(3)\prime}U_\alpha], \quad \text{(A.30)}
\end{aligned}$$

we can follow the analysis of $CV^*_{2,1}$ and show that by Assumptions A.4(ii) and A.6(i)

$$\begin{aligned}
CV^*_{2,3} &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(y_{it}-\hat{y}^{(3)}_{it})-\rho(y_{i,t-1}-\hat{y}^{(3)}_{i,t-1})\right]^2 \\
&= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(1-\rho)\,\alpha_i+v_{it}-\breve{x}^{(3)\prime}_{it}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}U_\alpha)\right]^2+o_P(1) \\
&= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(1-\rho)\,\alpha_i-\breve{x}^{(3)\prime}_{it}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_\alpha\underline{\alpha}\right]^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}v^2_{it}+o_P(1) \\
&\xrightarrow{P} c^*_{\alpha,X_\lambda}+\bar{\sigma}^2_v,
\end{aligned}$$

It follows that

$$CV^*_{2,3}-CV^*_{2,2}\xrightarrow{P} c^*_{\alpha,X_\lambda}>0. \quad \text{(A.31)}$$

To study $CV^*_{2,4}$, noting that

$$\begin{aligned}
& (y_{it}-\hat{y}^{(4)}_{it})-\rho(y_{i,t-1}-\hat{y}^{(4)}_{i,t-1}) \\
= & \frac{1}{1-h^{(4)}_{it}}\left[u_{it}-x^{(4)\prime}_{it}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right]-\frac{\rho}{1-h^{(4)}_{i,t-1}}\left[u_{it-1}-x^{(4)\prime}_{i,t-1}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right] \\
= & c_{it,4}\left[v_{it}-\breve{x}^{(4)\prime}_{it}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right]+\rho\varkappa_{it,4}\left[u_{i,t-1}-x^{(4)\prime}_{i,t-1}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right], \quad \text{(A.32)}
\end{aligned}$$

we have

$$\begin{aligned}
& CV^*_{2,4}-CV^*_{2,2} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c^2_{it,4}\left(v_{it}-\breve{x}^{(4)\prime}_{it}(X^{(4)\prime}X^{(4)})^{-1}X^{(4)\prime}U\right)^2-c^2_{it,2}\left(v_{it}-\breve{x}^{(2)\prime}_{it}(X^{(2)\prime}X^{(2)})^{-1}X^{(2)\prime}U\right)^2\right] \\
& +\frac{\rho^2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varkappa^2_{it,4}\left(u_{i,t-1}-x^{(4)\prime}_{i,t-1}(X^{(4)\prime}X^{(4)})^{-1}X^{(4)\prime}U\right)^2-\varkappa^2_{it,2}\left(u_{i,t-1}-x^{(2)\prime}_{i,t-1}(X^{(2)\prime}X^{(2)})^{-1}X^{(2)\prime}U\right)^2\right] \\
& +\frac{2\rho}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}\varkappa_{it,4}\left(v_{it}-\breve{x}^{(4)\prime}_{it}(X^{(4)\prime}X^{(4)})^{-1}X^{(4)\prime}U\right)\left(u_{i,t-1}-x^{(4)\prime}_{i,t-1}(X^{(4)\prime}X^{(4)})^{-1}X^{(4)\prime}U\right)\right. \\
& \left.-c_{it,2}\varkappa_{it,2}\left(v_{it}-\breve{x}^{(2)\prime}_{it}(X^{(2)\prime}X^{(2)})^{-1}X^{(2)\prime}U\right)\left(u_{i,t-1}-x^{(2)\prime}_{i,t-1}(X^{(2)\prime}X^{(2)})^{-1}X^{(2)\prime}U\right)\right] \\
\equiv & D_{2,4}(1)+\rho^2D_{2,4}(2)-2\rho D_{2,4}(3)\,,\ \text{say.}
\end{aligned}$$

For $D_{2,4}(1)$, we further make the following decomposition:

$$\begin{aligned}
D_{2,4}(1) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,4}^{2}-c_{it,2}^{2}\right)v_{it}^{2} \\
&\quad +\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[c_{it,4}^{2}\left(\breve{x}_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U\right)^{2}-c_{it,2}^{2}\left(\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^{2}\right] \\
&\quad -\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\left[c_{it,4}^{2}\breve{x}_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)\prime}U-c_{it,2}^{2}\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right] \\
&\equiv D_{2,4}(1,1)+D_{2,4}(1,2)-2D_{2,4}(1,3),\ \text{say.}
\end{aligned}$$

Following the analysis of $CV_{1,4}^{*}-CV_{1,1}^{*}$ in Case 1 and that of $CV_{2,4}-CV_{2,1}$ in the proof of Theorem 2.1, and applying Lemmas A.8(ii) and (iv) and A.9, (A.10) and (A.24), we can readily show that

$$\begin{aligned}
D_{2,4}(1,1) &= N^{-1}\frac{2}{NT_{1}}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^{2}+o_{P}\left(N^{-1}\right), \\
D_{2,4}(1,2) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[\left(\breve{x}_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)}U\right)^{2}-\left(\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}U\right)^{2}\right]+o_{P}\left(N^{-1}\right) \\
&= \frac{1}{T_{1}}\sum_{t=2}^{T}\left[(1-\rho L)\,\bar{u}_{\cdot t}\right]^{2}+o_{P}\left(N^{-1}\right),\ \text{and} \\
D_{2,4}(1,3) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}\left[\breve{x}_{it}^{(4)\prime}\left(X^{(4)\prime}X^{(4)}\right)^{-1}X^{(4)}U-\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)}U\right]+o_{P}\left(N^{-1}\right) \\
&= \frac{1}{NT_{1}}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}\left[(1-\rho L)\,\bar{u}_{\cdot t}\right]+o_{P}\left(N^{-1}\right).
\end{aligned}$$

It follows that $N\cdot D_{2,4}(1)=\frac{2}{NT_{1}}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^{2}-\frac{N}{T_{1}}\sum_{t=2}^{T}v_{\cdot t}^{2}+\frac{N}{T_{1}}\sum_{t=2}^{T}\left[v_{\cdot t}-(1-\rho L)\,\bar{u}_{\cdot t}\right]^{2}+o_{P}(1)$. Similarly, we can show that $D_{2,4}(\ell)=o_{P}\left(N^{-1}\right)$ for $\ell=2,3$. Consequently, we have by Assumptions A.4(ii) and A.5(ii)

$$N\left[CV_{2,4}^{*}-CV_{2,2}^{*}\right]=\frac{2}{NT_{1}}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^{2}-\frac{N}{T_{1}}\sum_{t=2}^{T}v_{\cdot t}^{2}+o_{P}(1)\xrightarrow{P}2\bar{\sigma}_{v}^{2}-\bar{\sigma}_{v2}^{2}. \tag{A.33}$$

By (A.29), (A.31), and (A.33), we have $P\left(CV_{2,2}^{*}<CV_{2,m}^{*}\right)\to 1$ as $(N,T)\to\infty$ for $m=1,3,4$ provided $\bar{\sigma}_{v2}^{2}<2\bar{\sigma}_{v}^{2}$.

Case 3: Model 3 is the true model. This case parallels Case 2 and we can follow the analysis in Case 2 and show that $P\left(CV_{3,3}^{*}<CV_{3,m}^{*}\right)\to 1$ for $m=1,2,4$. The details are omitted for brevity.

Case 4: Model 4 is the true model. In this case, Models 1-3 are underfitted and we will show that $P\left(CV_{4,4}^{*}<CV_{4,m}^{*}\right)\to 1$ for $m=1,2,3$. Let $u_{\lambda,it}$, $u_{\alpha\lambda,it}$, U_{λ}, and $U_{\alpha\lambda}$ be as defined in the proof of Theorem 2.1. Following the steps to obtain (A.23), now we can show that

$$\begin{aligned}
&(y_{it}-\hat{y}_{it}^{(1)})-\rho(y_{i,t-1}-\hat{y}_{i,t-1}^{(1)}) \\
&= \frac{1}{1-h_{it}^{(1)}}[u_{\alpha\lambda,it}-x_{it}^{\prime}(X^{\prime}X)^{-1}X^{\prime}U_{\alpha\lambda}]-\frac{\rho}{1-h_{i,t-1}^{(1)}}[u_{\alpha\lambda,it-1}-x_{i,t-1}^{\prime}(X^{\prime}X)^{-1}X^{\prime}U_{\alpha\lambda}] \\
&= c_{it,1}[(1-\rho)\,\alpha_{i}+(1-\rho L)\,\lambda_{t}+v_{it}-\breve{x}_{it}^{\prime}(X^{\prime}X)^{-1}X^{\prime}U_{\alpha\lambda}]+\rho\varkappa_{it,1}[u_{\alpha\lambda,it-1}-x_{i,t-1}^{\prime}(X^{\prime}X)^{-1}X^{\prime}U_{\alpha\lambda}]
\end{aligned} \tag{A.34}$$

where L denotes the lag operator. As in Case 2, we can show that by Assumptions A.4(iv)-(v) and A.6(iii),

$$\begin{aligned} CV_{4,1}^{*} &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(1-\rho)\,\alpha_i+(1-\rho L)\,\lambda_t-\breve{x}_{it}'\left(X'X\right)^{-1}X'\left(D_{\alpha}\underline{\alpha}+D_{\lambda}\underline{\lambda}\right)\right]^2+\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^2+o_P\left(1\right) \\ &\overset{P}{\to} c_{\alpha\lambda,X}^{*}+\bar{\sigma}_v^2. \end{aligned}$$

Similarly, we have

$$\begin{aligned} CV_{4,2}^{*} &= \frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}\left[(1-\rho L)\,\lambda_t-\breve{x}_{it}^{(2)\prime}\left(X^{(2)\prime}X^{(2)}\right)^{-1}X^{(2)\prime}D_{\lambda}\underline{\lambda}\right]^2+\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^2+o_P\left(1\right) \\ &\overset{P}{\to} c_{\lambda,X_{\alpha}}^{*}+\bar{\sigma}_v^2, \\ CV_{4,3}^{*} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(1-\rho)\,\alpha_i-\breve{x}_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_{\alpha}\underline{\alpha}\right]^2+\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^2+o_P\left(1\right) \\ &\overset{P}{\to} c_{\alpha,X_{\lambda}}^{*}+\bar{\sigma}_v^2, \end{aligned}$$

and $CV_{4,4}^{*}=\frac{1}{NT_1}\sum_{i=1}^{N}\sum_{t=2}^{T}v_{it}^2+o_P\left(1\right)\overset{P}{\to}\bar{\sigma}_v^2$. Then $P\left(CV_{4,4}^{*}<CV_{4,m}^{*}\right)\to 1$ as $(N,T)\to\infty$ for $m=1,2,3$. ■

REFERENCES

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction, *Technometrics* 16, 125-127.

Andrews, D. W. K. (1991). Asymptotic optimality of generalized C_L, cross-validation, and generalized cross-validation in regression with heteroskedastic errors, *Journal of Econometrics* 47, 359-377.

Arellano, M. (2003). *Panel Data Econometrics.* Oxford University Press, Oxford.

Ayres, I. and J. J. Donohue (2003). Shooting down the “More Guns, less Crime” hypothesis, *Stanford Law Review* 55, 1193-1312.

Bai, J. (2009). Panel data models with interactive fixed effects, *Econometrica* 77, 1229-1279.

Balazsi, L., L. Matyas, and T. Wansbeek (2016). Fixed effects models. In L. Matyas (eds), *The Econometrics of Multi-dimensional Panels,* forthcoming. Springer Verlag.

Baltagi, B. H. (2006). Estimating an economic model of crime using panel data from North Carolina, *Journal of Applied Econometrics* 21, 543-547

Baltagi, B. H. (2013). *Econometric Analysis of Panel Data.* 5th edition. John Wiley & Sons, West Sussex.

Baltagi, B. H., Y. Chang, and Q. Li (1992). Monte Carlo results on several new and existing tests for the error component model, *Journal of Econometrics* 54, 95-120.

Bera, A., W. Sosa-Escudero, and M. Yoon (2001). Tests for the error-component model in the presence of local misspecification, *Journal of Econometrics* 101, 1-23.

Bernstein, D. S. (2005). *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, Princeton.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

Breusch, T. S. and A. R. Pagan (1980). The Lagrange multiplier test and its applications to model specification in econometrics, *Reviews of Economic Studies* 47, 239-253.

Cornwell, C. and W. N. Trumbull (1994). Estimating the economic model of crime with panel data, *Review of Economics and Statistics* 76, 360-366.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models, *Review of Economic Studies* 82, 991-1030.

Edwards, S (1996). Why are Latin America's savings rates so low? An international comparative analysis, *Journal of Development Economics* 51, 5-44.

Geisser, S. (1974). The predictive sample reuse method with applications, *Journal of the American Statistical Association* 70, 320-328.

Gourieroux, C., A. Holly, and A. Monfort (1982). Likelihood ratio test, Wald test, and Kuhn–Tucker test in linear models with inequality constraints on the regression parameters, *Econometrica,* 50, 63-80.

Greene, W. H. (2008). *Econometric Analysis*. 6th edition. Pearson/Prentice Hall, New Jersey.

Hansen, B. E. and J. Racine (2012). Jackknife model averaging, *Journal of Econometrics* 167, 38-46.

Honda, Y. (1985). Testing the error components model with non-normal disturbances, *Review of Economic Studies* 52, 681-690.

Hsiao, C. (2014). *Analysis of Panel Data*. 3rd edition, Cambridge University Press, Cambridge.

Li, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: discrete index set, *Annals of Statistics* 15, 958-975.

Lu, X. and L. Su (2015). Jackknife model averaging for quantile regressions, *Journal of Econometrics*, 188, 40-58.

Moulton, B. and W. Randolph (1989). Alternative tests of the error components model, *Econometrica* 57, 685-693.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* 74, 967-1012.

Romano, J., A. Shaikh, and W. Michael (2010). Multiple testing. In S. N. Durlauf and L. E. Blume (eds.), *The New Palgrave Dictionary of Economics*, Online ed., UK: Palgrave Macmillan.

Stock, J. and M. Watson (2012). *Introduction to Econometrics*. 3rd edition, Pearson, Boston.

Stone, C. (1974). Cross-validatory choice and assessment of statistical predictions, *Journal of the Royal Statistical Society*, *Series B* 36, 111-147.

Su, L., Z. Shi, and P. C. B. Phillips (2016). Identifying latent structures in panel data, *Econometrica* 84, 2215-2264.

Wahba, G. and S. Wold (1975). A completely automatic french curve: fitting spline functions by cross-validation, *Communications in Statistics* 4, 1-17.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* 2nd edition. MIT Press, Massachusetts.

Wu, J. and G. Li (2014). Moment-based tests for individual and time effects in panel data models, *Journal of Econometrics* 178, 569-581.

Wu, J., J. Qin, and Q. Ding (2015). A moment-based test for individual effects in the error component model with incomplete panels, *Statistics and Probability Letters* 104, 153-162.

Wu, J. and L. X. Zhu (2012). Estimation of and testing for random effects in dynamic panel data models, *Test* 21, 477-497.

Table 1A: Frequency of the model selected: static panels, $\rho = 0$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.897**	0.059	0.038	0.006	0.006	**0.913**	0	0.081
	N=10 T=50	**0.971**	0.028	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.962**	0	0.038	0	0	**0.932**	0	0.068
	N=50 T=50	**1**	0	0	0	0	**0.999**	0	0.001
BIC	N=10 T=10	**1**	0	0	0	0.136	**0.864**	0	0
	N=10 T=50	**1**	0	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	0.411	**0.392**	0.197	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.649**	0.134	0.157	0.060	0	**0.739**	0.001	0.260
	N=10 T=50	**0.939**	0.060	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.934**	0.001	0.065	0	0	**0.895**	0	0.105
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.931**	0.039	0.030	0	0.007	**0.961**	0	0.032
	N=10 T=50	**0.974**	0.026	0	0	0	**1**	0	0
	N=50 T=10	**0.963**	0	0.037	0	0	**0.965**	0	0.035
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV^*	N=10 T=10	**0.808**	0.143	0.041	0.008	0.005	**0.944**	0.001	0.050
	N=10 T=50	**0.959**	0.040	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.938**	0.013	0.049	0	0	**0.944**	0	0.056
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV^{**}	N=10 T=10	**0.877**	0.076	0.042	0.005	0.044	**0.903**	0.014	0.039
	N=10 T=50	**0.965**	0.034	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.953**	0	0.047	0	0	**0.945**	0.001	0.054
	N=50 T=50	**1**	0	0.001	0	0	**0.999**	0	0.001

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.008	0.001	**0.898**	0.093	0.003	0.009	0.006	**0.982**
	N=10 T=50	0	0	**0.946**	0.054	0	0	0	**1**
	N=50 T=10	0	0	**0.999**	0.001	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.200	0	**0.800**	0	0.740	0.029	0.011	**0.220**
	N=10 T=50	0.349	0.278	**0.373**	0	0.404	0.561	0	**0.035**
	N=50 T=10	0	0	**1**	0	0.542	0	0.413	**0.045**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.002	0	**0.758**	0.240	0	0.002	0.001	**0.997**
	N=10 T=50	0	0	**0.902**	0.098	0	0	0	**1**
	N=50 T=10	0	0	**0.999**	0.001	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.010	0.001	**0.950**	0.039	0.007	0.019	0.010	**0.964**
	N=10 T=50	0	0	**0.976**	0.024	0	0	0	**1**
	N=50 T=10	0	0	**1**	0	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV^*	N=10 T=10	0.008	0.005	**0.856**	0.131	0.014	0.023	0.006	**0.957**
	N=10 T=50	0	0	**0.959**	0.041	0	0	0	**1**
	N=50 T=10	0	0	**0.989**	0.011	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV^{**}	N=10 T=10	0.020	0.003	**0.904**	0.073	0.021	0.026	0.122	**0.831**
	N=10 T=50	0	0	**0.965**	0.035	0	0	0	**1**
	N=50 T=10	0	0	**1**	0	0	0	0.005	**0.995**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 1B: Frequency of the model selected: static panels, $\rho = 1/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.607**	0.318	0.034	0.041	0.006	**0.893**	0	0.101
	N=10 T=50	**0.718**	0.282	0	0	0	**1**	0	0
	N=50 T=10	**0.760**	0.194	0.028	0.018	0	**0.915**	0	0.085
	N=50 T=50	**0.855**	0.145	0	0	0	**1**	0	0
BIC	N=10 T=10	**0.997**	0.003	0	0	0.106	**0.894**	0	0
	N=10 T=50	**1**	0	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	0.313	**0.590**	0.097	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.337**	0.428	0.077	0.158	0	**0.742**	0.001	0.257
	N=10 T=50	**0.646**	0.35	0.004	0	0	**0.995**	0	0.005
	N=50 T=10	**0.585**	0.334	0.039	0.042	0	**0.889**	0	0.111
	N=50 T=50	**0.883**	0.117	0	0	0	**1**	0	0
CV	N=10 T=10	**0.694**	0.268	0.027	0.011	0.007	**0.949**	0	0.044
	N=10 T=50	**0.731**	0.269	0	0	0	**1**	0	0
	N=50 T=10	**0.840**	0.125	0.028	0.007	0	**0.954**	0	0.046
	N=50 T=50	**0.870**	0.130	0	0	0	**1**	0	0
CV^*	N=10 T=10	**0.738**	0.210	0.036	0.016	0.013	**0.935**	0.002	0.050
	N=10 T=50	**0.950**	0.049	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.898**	0.052	0.047	0.003	0	**0.946**	0	0.054
	N=50 T=50	**0.999**	0.001	0	0	0	**0.999**	0	0.001
CV^{**}	N=10 T=10	**0.858**	0.093	0.042	0.007	0.184	**0.743**	0.039	0.034
	N=10 T=50	**0.960**	0.039	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.952**	0.001	0.047	0	0.027	**0.867**	0.061	0.045
	N=50 T=50	**0.999**	0	0.001	0	0	**0.999**	0	0.001
True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.006	0.001	**0.575**	0.418	0.002	0.011	0.005	**0.982**
	N=10 T=50	0	0	**0.633**	0.367	0	0	0	**1**
	N=50 T=10	0	0	**0.765**	0.235	0	0	0	**1**
	N=50 T=50	0	0	**0.830**	0.170	0	0	0	**1**
BIC	N=10 T=10	0.238	0.003	**0.752**	0.007	0.648	0.051	0.011	**0.290**
	N=10 T=50	0.367	0.39	**0.243**	0	0.379	0.599	0	**0.022**
	N=50 T=10	0.002	0	**0.998**	0	0.505	0	0.333	**0.162**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.001	0.001	**0.372**	0.626	0	0.002	0.001	**0.997**
	N=10 T=50	0	0	**0.542**	0.458	0	0	0	**1**
	N=50 T=10	0	0	**0.593**	0.407	0	0	0	**1**
	N=50 T=50	0	0	**0.870**	0.130	0	0	0	**1**
CV	N=10 T=10	0.012	0.005	**0.714**	0.269	0.003	0.021	0.008	**0.968**
	N=10 T=50	0	0	**0.724**	0.276	0	0	0	**1**
	N=50 T=10	0	0	**0.871**	0.129	0	0	0	**1**
	N=50 T=50	0	0	**0.872**	0.128	0	0	0	**1**
CV^*	N=10 T=10	0.011	0.005	**0.773**	0.211	0.01	0.018	0.02	**0.952**
	N=10 T=50	0	0	**0.951**	0.049	0	0	0	**1**
	N=50 T=10	0	0	**0.950**	0.050	0	0	0	**1**
	N=50 T=50	0	0	**0.999**	0.001	0	0	0	**1**
CV^{**}	N=10 T=10	0.016	0.006	**0.888**	0.090	0.023	0.016	0.348	**0.613**
	N=10 T=50	0	0	**0.964**	0.036	0	0	0	**1**
	N=50 T=10	0	0	**1**	0	0	0	0.184	**0.816**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 1C: Frequency of the model selected: static panels, $\rho = 1/3$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.472**	0.438	0.031	0.059	0.006	**0.887**	0	0.107
	N=10 T=50	**0.566**	0.433	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.446**	0.491	0.019	0.044	0	**0.914**	0	0.086
	N=50 T=50	**0.534**	0.466	0	0	0	**1**	0	0
BIC	N=10 T=10	**0.987**	0.013	0	0	0.104	**0.896**	0	0
	N=10 T=50	**0.998**	0.002	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	0.283	**0.646**	0.071	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.239**	0.516	0.048	0.197	0.001	**0.743**	0.002	0.254
	N=10 T=50	**0.474**	0.519	0.005	0.002	0	**0.993**	0	0.007
	N=50 T=10	**0.277**	0.619	0.019	0.085	0	**0.885**	0	0.115
	N=50 T=50	**0.601**	0.399	0	0	0	**1**	0	0
CV	N=10 T=10	**0.541**	0.415	0.022	0.022	0.007	**0.940**	0	0.053
	N=10 T=50	**0.578**	0.422	0	0	0	**1**	0	0
	N=50 T=10	**0.548**	0.412	0.021	0.019	0	**0.951**	0	0.049
	N=50 T=50	**0.557**	0.443	0	0	0	**1**	0	0
CV^*	N=10 T=10	**0.694**	0.251	0.037	0.018	0.02	**0.925**	0.003	0.052
	N=10 T=50	**0.945**	0.054	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.867**	0.083	0.043	0.007	0	**0.946**	0	0.054
	N=50 T=50	**0.999**	0.001	0	0	0	**0.999**	0	0.001
CV^{**}	N=10 T=10	**0.842**	0.105	0.044	0.009	0.254	**0.668**	0.052	0.026
	N=10 T=50	**0.960**	0.039	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.951**	0.002	0.047	0	0.091	**0.751**	0.125	0.033
	N=50 T=50	**0.999**	0	0.001	0	0	**0.999**	0	0.001

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.009	0.003	**0.441**	0.547	0.002	0.011	0.005	**0.982**
	N=10 T=50	0	0	**0.484**	0.516	0	0	0	**1**
	N=50 T=10	0	0	**0.441**	0.559	0	0	0	**1**
	N=50 T=50	0	0	**0.494**	0.506	0	0	0	**1**
BIC	N=10 T=10	0.273	0.011	**0.703**	0.013	0.611	0.064	0.008	**0.317**
	N=10 T=50	0.382	0.450	**0.168**	0	0.373	0.611	0	**0.016**
	N=50 T=10	0.002	0	**0.998**	0	0.49	0	0.287	**0.223**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.001	0.001	**0.245**	0.753	0	0.002	0.002	**0.996**
	N=10 T=50	0	0	**0.409**	0.591	0	0	0	**1**
	N=50 T=10	0	0	**0.276**	0.724	0	0	0	**1**
	N=50 T=50	0	0	**0.560**	0.440	0	0	0	**1**
CV	N=10 T=10	0.011	0.009	**0.568**	0.412	0.003	0.022	0.009	**0.966**
	N=10 T=50	0	0	**0.575**	0.425	0	0	0	**1**
	N=50 T=10	0	0	**0.580**	0.420	0	0	0	**1**
	N=50 T=50	0	0	**0.563**	0.437	0	0	0	**1**
CV^*	N=10 T=10	0.01	0.004	**0.730**	0.256	0.009	0.015	0.022	**0.954**
	N=10 T=50	0	0	**0.942**	0.058	0	0	0	**1**
	N=50 T=10	0	0	**0.913**	0.087	0	0	0	**1**
	N=50 T=50	0	0	**0.999**	0.001	0	0	0	**1**
CV^{**}	N=10 T=10	0.014	0.005	**0.880**	0.101	0.016	0.014	0.449	**0.521**
	N=10 T=50	0	0	**0.963**	0.037	0	0	0	**1**
	N=50 T=10	0	0	**0.998**	0.002	0	0	0.344	**0.656**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 1D: Frequency of the model selected: static panels, $\rho = 1/2$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.209**	0.680	0.014	0.097	0.005	**0.871**	0	0.124
	N=10 T=50	**0.273**	0.723	0	0.004	0	**0.996**	0	0.004
	N=50 T=10	**0.037**	0.873	0	0.090	0	**0.909**	0	0.091
	N=50 T=50	**0.039**	0.960	0	0.001	0	**0.999**	0	0.001
BIC	N=10 T=10	**0.888**	0.112	0	0	0.093	**0.905**	0	0.002
	N=10 T=50	**0.981**	0.019	0	0	0.001	**0.999**	0	0
	N=50 T=10	**1**	0	0	0	0.204	**0.780**	0.016	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.083**	0.668	0.023	0.226	0.001	**0.744**	0.001	0.254
	N=10 T=50	**0.217**	0.771	0.002	0.010	0	**0.988**	0	0.012
	N=50 T=10	**0.008**	0.867	0.001	0.124	0	**0.875**	0	0.125
	N=50 T=50	**0.055**	0.944	0	0.001	0	**0.999**	0	0.001
CV	N=10 T=10	**0.248**	0.692	0.012	0.048	0.006	**0.932**	0	0.062
	N=10 T=50	**0.282**	0.715	0.001	0.002	0	**0.997**	0	0.003
	N=50 T=10	**0.057**	0.883	0	0.06	0	**0.938**	0	0.062
	N=50 T=50	**0.044**	0.955	0	0.001	0	**0.999**	0	0.001
CV^*	N=10 T=10	**0.579**	0.362	0.032	0.027	0.028	**0.914**	0.003	0.055
	N=10 T=50	**0.908**	0.091	0.001	0	0	**0.999**	0	0.001
	N=50 T=10	**0.692**	0.261	0.034	0.013	0	**0.946**	0	0.054
	N=50 T=50	**0.998**	0.002	0	0	0	**0.999**	0	0.001
CV^{**}	N=10 T=10	**0.809**	0.141	0.039	0.011	0.42	**0.505**	0.054	0.021
	N=10 T=50	**0.950**	0.049	0.001	0	0.002	**0.997**	0	0.001
	N=50 T=10	**0.945**	0.009	0.046	0	0.312	**0.444**	0.229	0.015
	N=50 T=50	**0.998**	0.001	0.001	0	0	**0.999**	0	0.001

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.003	0.013	**0.176**	0.808	0.001	0.015	0.004	**0.980**
	N=10 T=50	0	0	**0.220**	0.780	0	0	0	**1**
	N=50 T=10	0	0	**0.030**	0.970	0	0	0	**1**
	N=50 T=50	0	0	**0.034**	0.966	0	0	0	**1**
BIC	N=10 T=10	0.362	0.043	**0.493**	0.102	0.510	0.122	0.007	**0.361**
	N=10 T=50	0.390	0.578	**0.031**	0.001	0.371	0.624	0	**0.005**
	N=50 T=10	0.005	0	**0.995**	0	0.425	0	0.159	**0.416**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0	0.006	**0.091**	0.903	0	0.006	0.001	**0.993**
	N=10 T=50	0	0	**0.178**	0.822	0	0	0	**1**
	N=50 T=10	0	0	**0.009**	0.991	0	0	0	**1**
	N=50 T=50	0	0	**0.044**	0.956	0	0	0	**1**
CV	N=10 T=10	0.005	0.021	**0.280**	0.694	0.002	0.026	0.007	**0.965**
	N=10 T=50	0	0.001	**0.286**	0.713	0	0.001	0	**0.999**
	N=50 T=10	0	0	**0.058**	0.942	0	0	0	**1**
	N=50 T=50	0	0	**0.046**	0.954	0	0	0	**1**
CV^*	N=10 T=10	0.01	0.005	**0.612**	0.373	0.007	0.014	0.034	**0.945**
	N=10 T=50	0	0	**0.904**	0.096	0	0	0	**1**
	N=50 T=10	0	0	**0.735**	0.265	0	0	0	**1**
	N=50 T=50	0	0	**0.998**	0.002	0	0	0	**1**
CV^{**}	N=10 T=10	0.011	0.005	**0.842**	0.142	0.014	0.008	0.583	**0.395**
	N=10 T=50	0	0	**0.957**	0.043	0	0	0.024	**0.976**
	N=50 T=10	0	0	**0.993**	0.007	0	0	0.694	**0.306**
	N=50 T=50	0	0	**0.999**	0.001	0	0	0	**1**

Table 1E: Frequency of the model selected: static panels, $\rho = 3/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.014**	0.842	0.003	0.141	0.001	**0.853**	0	0.146
	N=10 T=50	**0.019**	0.953	0	0.028	0	**0.972**	0	0.028
	N=50 T=10	**0**	0.873	0	0.127	0	**0.873**	0	0.127
	N=50 T=50	**0**	0.982	0	0.018	0	**0.982**	0	0.018
BIC	N=10 T=10	**0.242**	0.752	0	0.006	0.04	**0.954**	0	0.006
	N=10 T=50	**0.466**	0.534	0	0	0.002	**0.998**	0	0
	N=50 T=10	**0.467**	0.533	0	0	0.007	**0.993**	0	0
	N=50 T=50	**0.770**	0.230	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.003**	0.739	0.002	0.256	0	**0.742**	0	0.258
	N=10 T=50	**0.012**	0.942	0	0.046	0	**0.954**	0	0.046
	N=50 T=10	**0**	0.841	0	0.159	0	**0.841**	0	0.159
	N=50 T=50	**0**	0.983	0	0.017	0	**0.983**	0	0.017
CV	N=10 T=10	**0.021**	0.897	0.005	0.077	0.001	**0.918**	0	0.081
	N=10 T=50	**0.021**	0.962	0.001	0.016	0	**0.983**	0	0.017
	N=50 T=10	**0**	0.914	0	0.086	0	**0.914**	0	0.086
	N=50 T=50	**0**	0.983	0	0.017	0	**0.983**	0	0.017
CV^*	N=10 T=10	**0.302**	0.633	0.021	0.044	0.038	**0.896**	0.001	0.065
	N=10 T=50	**0.705**	0.294	0.001	0	0.009	**0.990**	0	0.001
	N=50 T=10	**0.153**	0.788	0.009	0.050	0	**0.941**	0	0.059
	N=50 T=50	**0.962**	0.038	0	0	0	**0.999**	0	0.001
CV^{**}	N=10 T=10	**0.660**	0.292	0.032	0.016	0.532	**0.412**	0.035	0.021
	N=10 T=50	**0.918**	0.081	0.001	0	0.320	**0.679**	0.001	0
	N=50 T=10	**0.867**	0.088	0.043	0.002	0.648	**0.231**	0.113	0.008
	N=50 T=50	**0.998**	0.001	0.001	0	0.157	**0.842**	0	0.001
True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.002	0.029	**0.013**	0.956	0	0.030	0.001	**0.969**
	N=10 T=50	0	0.012	**0.013**	0.975	0	0.012	0	**0.988**
	N=50 T=10	0	0	**0**	1	0	0	0	**1**
	N=50 T=50	0	0	**0**	1	0	0	0	**1**
BIC	N=10 T=10	0.192	0.356	**0.075**	0.377	0.190	0.381	0.005	**0.424**
	N=10 T=50	0.245	0.755	**0**	0	0.229	0.771	0	**0**
	N=50 T=10	0.016	0.002	**0.419**	0.563	0.058	0.005	0.002	**0.935**
	N=50 T=50	0.001	0.001	**0.743**	0.255	0.001	0.002	0	**0.997**
BIC_2	N=10 T=10	0	0.013	**0.005**	0.982	0	0.013	0	**0.987**
	N=10 T=50	0	0.005	**0.010**	0.985	0	0.005	0	**0.995**
	N=50 T=10	0	0	**0**	1	0	0	0	**1**
	N=50 T=50	0	0	**0**	1	0	0	0	**1**
CV	N=10 T=10	0.005	0.059	**0.025**	0.911	0.001	0.060	0.002	**0.937**
	N=10 T=50	0	0.021	**0.025**	0.954	0	0.021	0	**0.979**
	N=50 T=10	0	0	**0**	1	0	0	0	**1**
	N=50 T=50	0	0	**0**	1	0	0	0	**1**
CV^*	N=10 T=10	0.005	0.003	**0.317**	0.675	0.003	0.007	0.05	**0.940**
	N=10 T=50	0	0	**0.683**	0.317	0	0	0.003	**0.997**
	N=50 T=10	0	0	**0.169**	0.831	0	0	0	**1**
	N=50 T=50	0	0	**0.961**	0.039	0	0	0	**1**
CV^{**}	N=10 T=10	0.006	0	**0.732**	0.262	0.006	0	0.637	**0.357**
	N=10 T=50	0	0	**0.929**	0.071	0	0	0.524	**0.476**
	N=50 T=10	0	0	**0.911**	0.089	0	0	0.815	**0.185**
	N=50 T=50	0	0	**0.999**	0.001	0	0	0.509	**0.491**

Table 2A: Frequency of the model selected: dynamic panels without exogenous regressors, $\beta = 1/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.847**	0.095	0.041	0.017	0.005	**0.909**	0	0.086
	N=10 T=50	**0.954**	0.046	0	0	0	**1**	0	0
	N=50 T=10	**0.958**	0.004	0.038	0	0	**0.928**	0	0.072
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**1**	0	0	0	0.406	**0.594**	0	0
	N=10 T=50	**1**	0	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	1	**0**	0	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.572**	0.221	0.116	0.091	0	**0.764**	0	0.236
	N=10 T=50	**0.925**	0.074	0.001	0	0	**0.997**	0	0.003
	N=50 T=10	**0.917**	0.018	0.061	0.004	0	**0.894**	0	0.106
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.887**	0.075	0.032	0.006	0.007	**0.954**	0	0.039
	N=10 T=50	**0.955**	0.045	0	0	0	**1**	0	0
	N=50 T=10	**0.962**	0.002	0.036	0	0	**0.960**	0	0.04
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.005	0.001	**0.845**	0.149	0	0.005	0.001	**0.994**
	N=10 T=50	0	0	**0.924**	0.076	0	0	0	**1**
	N=50 T=10	0	0	**0.994**	0.006	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.114	0	**0.886**	0	0.072	0.046	0.313	**0.569**
	N=10 T=50	0.083	0	**0.917**	0	0.001	0.075	0	**0.924**
	N=50 T=10	0.002	0	**0.998**	0	0.003	0	0.997	**0**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.001	0.001	**0.654**	0.344	0	0.002	0	**0.998**
	N=10 T=50	0	0	**0.885**	0.115	0	0	0	**1**
	N=50 T=10	0	0	**0.974**	0.026	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.01	0.001	**0.912**	0.077	0	0.011	0.007	**0.982**
	N=10 T=50	0	0	**0.955**	0.045	0	0	0	**1**
	N=50 T=10	0	0	**0.999**	0.001	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 2B: Frequency of the model selected: dynamic panels without exogenous regressors, $\beta = 1/2$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.796**	0.143	0.038	0.023	0.008	**0.905**	0.001	0.086
	N=10 T=50	**0.948**	0.052	0	0	0	**1**	0	0
	N=50 T=10	**0.942**	0.017	0.038	0.003	0	**0.930**	0	0.07
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**1**	0	0	0	0.759	**0.241**	0	0
	N=10 T=50	**1**	0	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	1	**0**	0	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.508**	0.281	0.093	0.118	0	**0.759**	0.001	0.240
	N=10 T=50	**0.920**	0.078	0.002	0	0	**0.997**	0	0.003
	N=50 T=10	**0.882**	0.050	0.060	0.008	0	**0.897**	0	0.103
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.844**	0.123	0.026	0.007	0.012	**0.951**	0.001	0.036
	N=10 T=50	**0.953**	0.047	0	0	0	**1**	0	0
	N=50 T=10	**0.951**	0.012	0.037	0	0	**0.957**	0	0.043
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.006	0.001	**0.778**	0.215	0	0.005	0.008	**0.987**
	N=10 T=50	0	0	**0.915**	0.085	0	0	0	**1**
	N=50 T=10	0	0	**0.977**	0.023	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.119	0	**0.880**	0.001	0.1	0.023	0.635	**0.242**
	N=10 T=50	0.085	0	**0.915**	0	0.003	0.071	0	**0.926**
	N=50 T=10	0.002	0	**0.998**	0	0.003	0	0.997	**0**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.001	0.001	**0.551**	0.447	0	0.002	0.001	**0.997**
	N=10 T=50	0	0	**0.874**	0.126	0	0	0	**1**
	N=50 T=10	0	0	**0.936**	0.064	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.009	0.001	**0.878**	0.112	0.001	0.01	0.025	**0.964**
	N=10 T=50	0	0	**0.952**	0.048	0	0	0	**1**
	N=50 T=10	0	0	**0.988**	0.012	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 2C: Frequency of the model selected: dynamic panels without exogenous regressors, $\beta = 3/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.633**	0.303	0.032	0.032	0.048	**0.872**	0.002	0.078
	N=10 T=50	**0.924**	0.076	0	0	0	**1**	0	0
	N=50 T=10	**0.834**	0.122	0.029	0.015	0	**0.925**	0	0.075
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**0.999**	0.001	0	0	0.950	**0.050**	0	0
	N=10 T=50	**1**	0	0	0	0.080	**0.920**	0	0
	N=50 T=10	**1**	0	0	0	1	**0**	0	0
	N=50 T=50	**1**	0	0	0	0.905	**0.095**	0	0
BIC_2	N=10 T=10	**0.334**	0.446	0.053	0.167	0.005	**0.754**	0.002	0.239
	N=10 T=50	**0.890**	0.108	0.002	0	0	**0.998**	0	0.002
	N=50 T=10	**0.663**	0.260	0.040	0.037	0	**0.898**	0	0.102
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.705**	0.262	0.023	0.010	0.082	**0.880**	0.002	0.036
	N=10 T=50	**0.927**	0.073	0	0	0	**1**	0	0
	N=50 T=10	**0.891**	0.071	0.030	0.008	0.007	**0.952**	0.001	0.04
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.005	0.001	**0.609**	0.385	0	0.004	0.050	**0.946**
	N=10 T=50	0	0	**0.885**	0.115	0	0	0	**1**
	N=50 T=10	0	0	**0.845**	0.155	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.121	0	**0.874**	0.005	0.092	0.007	0.847	**0.054**
	N=10 T=50	0.087	0	**0.913**	0	0.027	0.052	0.077	**0.844**
	N=50 T=10	0.002	0	**0.998**	0	0.004	0	0.996	**0**
	N=50 T=50	0	0	**1**	0	0	0	0.896	**0.104**
BIC_2	N=10 T=10	0.001	0.001	**0.350**	0.648	0	0.002	0.006	**0.992**
	N=10 T=50	0	0	**0.835**	0.165	0	0	0	**1**
	N=50 T=10	0	0	**0.675**	0.325	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.007	0.002	**0.762**	0.229	0	0.008	0.139	**0.853**
	N=10 T=50	0	0	**0.927**	0.073	0	0	0	**1**
	N=50 T=10	0	0	**0.923**	0.077	0	0	0.010	**0.990**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 3A: Frequency of the model selected: dynamic panels with exogenous regressors, $\beta = 1/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.809**	0.112	0.046	0.033	0.009	**0.877**	0.006	0.108
	N=10 T=50	**0.940**	0.060	0	0	0	**1**	0	0
	N=50 T=10	**0.954**	0.002	0.044	0	0	**0.914**	0	0.086
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**1**	0	0	0	0.676	**0.324**	0	0
	N=10 T=50	**1**	0	0	0	0	**1**	0	0
	N=50 T=10	**1**	0	0	0	0.996	**0**	0.004	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.552**	0.208	0.115	0.125	0.001	**0.714**	0.003	0.282
	N=10 T=50	**0.919**	0.080	0.001	0	0	**0.997**	0	0.003
	N=50 T=10	**0.912**	0.013	0.074	0.001	0	**0.868**	0	0.132
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.913**	0.057	0.023	0.007	0.030	**0.931**	0.006	0.033
	N=10 T=50	**0.951**	0.049	0	0	0	**1**	0	0
	N=50 T=10	**0.963**	0	0.037	0	0	**0.957**	0	0.043
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.004	0.002	**0.799**	0.195	0.006	0.010	0.024	**0.960**
	N=10 T=50	0	0	**0.911**	0.089	0	0	0	**1**
	N=50 T=10	0	0	**0.997**	0.003	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.225	0.001	**0.774**	0	0.712	0.011	0.168	**0.109**
	N=10 T=50	0.384	0.283	**0.333**	0	0.794	0.180	0	**0.026**
	N=50 T=10	0.002	0	**0.998**	0	0.024	0	0.976	**0**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC_2	N=10 T=10	0.001	0.001	**0.617**	0.381	0	0.002	0.003	**0.995**
	N=10 T=50	0	0	**0.870**	0.13	0	0	0	**1**
	N=50 T=10	0	0	**0.979**	0.021	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.012	0.007	**0.919**	0.062	0.026	0.023	0.093	**0.858**
	N=10 T=50	0	0	**0.954**	0.046	0	0	0	**1**
	N=50 T=10	0	0	**1**	0	0	0	0	**1**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 3B: Frequency of the model selected: dynamic panels with exogenous regressors, $\beta = 1/2$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.770**	0.143	0.045	0.042	0.035	**0.849**	0.010	0.106
	N=10 T=50	**0.940**	0.060	0	0	0	**1**	0	0
	N=50 T=10	**0.946**	0.010	0.042	0.002	0.001	**0.903**	0	0.096
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**1**	0	0	0	0.869	**0.131**	0	0
	N=10 T=50	**1**	0	0	0	0.005	**0.995**	0	0
	N=50 T=10	**1**	0	0	0	1	**0**	0	0
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC_2	N=10 T=10	**0.503**	0.257	0.099	0.141	0.002	**0.711**	0.007	0.280
	N=10 T=50	**0.911**	0.088	0.001	0	0	**0.998**	0	0.002
	N=50 T=10	**0.886**	0.034	0.070	0.01	0	**0.858**	0	0.142
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.874**	0.096	0.021	0.009	0.08	**0.882**	0.007	0.031
	N=10 T=50	**0.947**	0.053	0	0	0	**1**	0	0
	N=50 T=10	**0.961**	0.003	0.036	0	0.002	**0.954**	0	0.044
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.003	0.002	**0.759**	0.236	0.006	0.007	0.061	**0.926**
	N=10 T=50	0	0	**0.904**	0.096	0	0	0	**1**
	N=50 T=10	0	0	**0.986**	0.014	0	0	0.003	**0.997**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.237	0.001	**0.762**	0	0.629	0.009	0.291	**0.071**
	N=10 T=50	0.401	0.270	**0.329**	0	0.917	0.064	0	**0.019**
	N=50 T=10	0.002	0	**0.998**	0	0.009	0	0.991	**0**
	N=50 T=50	0	0	**1**	0	0	0	0.009	**0.991**
BIC_2	N=10 T=10	0.001	0.001	**0.545**	0.453	0	0.003	0.011	**0.986**
	N=10 T=50	0	0	**0.865**	0.135	0	0	0	**1**
	N=50 T=10	0	0	**0.948**	0.052	0	0	0.001	**0.999**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.014	0.007	**0.879**	0.100	0.027	0.018	0.180	**0.775**
	N=10 T=50	0	0	**0.953**	0.047	0	0	0	**1**
	N=50 T=10	0	0	**0.998**	0.002	0	0	0.011	**0.989**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 3C: Frequency of the model selected: dynamic panels with exogenous regressors, $\beta = 3/4$

True model		Model 1				Model 2			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	**0.644**	0.265	0.034	0.057	0.129	**0.756**	0.012	0.103
	N=10 T=50	**0.925**	0.075	0	0	0	**1**	0	0
	N=50 T=10	**0.884**	0.070	0.031	0.015	0.040	**0.848**	0.002	0.110
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
BIC	N=10 T=10	**1**	0	0	0	0.964	**0.036**	0	0
	N=10 T=50	**1**	0	0	0	0.38	**0.620**	0	0
	N=50 T=10	**1**	0	0	0	1	**0**	0	0
	N=50 T=50	**1**	0	0	0	1	**0**	0	0
BIC_2	N=10 T=10	**0.361**	0.389	0.068	0.182	0.022	**0.694**	0.016	0.268
	N=10 T=50	**0.885**	0.113	0.002	0	0	**0.998**	0	0.002
	N=50 T=10	**0.748**	0.156	0.054	0.042	0.012	**0.831**	0.001	0.156
	N=50 T=50	**1**	0	0	0	0	**1**	0	0
CV	N=10 T=10	**0.778**	0.187	0.023	0.012	0.243	**0.714**	0.013	0.030
	N=10 T=50	**0.929**	0.071	0	0	0	**1**	0	0
	N=50 T=10	**0.931**	0.034	0.031	0.004	0.079	**0.854**	0.012	0.055
	N=50 T=50	**1**	0	0	0	0	**1**	0	0

True model		Model 3				Model 4			
Selected model		M1	M2	M3	M4	M1	M2	M3	M4
AIC	N=10 T=10	0.002	0.004	**0.611**	0.383	0.006	0.007	0.143	**0.844**
	N=10 T=50	0	0	**0.882**	0.118	0	0	0	**1**
	N=50 T=10	0	0	**0.894**	0.106	0	0	0.048	**0.952**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
BIC	N=10 T=10	0.260	0.003	**0.736**	0.001	0.478	0.008	0.485	**0.029**
	N=10 T=50	0.442	0.241	**0.317**	0	0.982	0.007	0	**0.011**
	N=50 T=10	0.002	0	**0.998**	0	0.006	0	0.994	**0**
	N=50 T=50	0	0	**1**	0	0	0	1	**0**
BIC_2	N=10 T=10	0.001	0.003	**0.370**	0.626	0	0.004	0.037	**0.959**
	N=10 T=50	0	0	**0.835**	0.165	0	0	0	**1**
	N=50 T=10	0	0	**0.776**	0.224	0	0	0.012	**0.988**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**
CV	N=10 T=10	0.011	0.011	**0.808**	0.170	0.027	0.014	0.366	**0.593**
	N=10 T=50	0	0	**0.928**	0.072	0	0	0	**1**
	N=50 T=10	0	0	**0.961**	0.039	0	0	0.133	**0.867**
	N=50 T=50	0	0	**1**	0	0	0	0	**1**

Table 4A: Comparisons of MSEs: static panels, $\rho = 0$

True Model	Adopted Model	M1	M2	M3	M4
M1	N=10 T=10	**3.79**	6.32	5.83	13.16
	N=10 T=50	**0.68**	1.03	0.99	2.16
	N=50 T=10	**0.71**	1.23	1.02	2.31
	N=50 T=50	**0.14**	0.21	0.21	0.41
M2	N=10 T=10	145.62	**6.32**	295.26	13.16
	N=10 T=50	128.66	**1.03**	287.75	2.16
	N=50 T=10	130.75	**1.23**	259.24	2.31
	N=50 T=50	117.76	**0.21**	258.94	0.41
M3	N=10 T=10	102.06	245.89	**5.83**	13.16
	N=10 T=50	109.93	249.80	**0.99**	2.16
	N=50 T=10	94.97	235.31	**1.02**	2.31
	N=50 T=50	107.42	246.93	**0.21**	0.41
M4	N=10 T=10	440.18	245.89	295.26	**13.16**
	N=10 T=50	448.25	249.80	287.75	**2.16**
	N=50 T=10	422.87	235.31	259.24	**2.31**
	N=50 T=50	441.74	246.93	258.94	**0.41**

Note: Numbers in the main entries are 1000×MSEs of the estimates of β.

Table 4B: Comparisons of MSEs: dynamic panels without exogenous regressors, $\beta = 3/4$

		Non-bias correction				Bias correction			
True Model	Adopted Model	M1	M2	M3	M4	M1	M2	M3	M4
M1	N=10 T=10	**5.55**	57.34	5.40	58.74	**5.41**	25.20	5.52	28.49
	N=10 T=50	**0.84**	2.48	0.92	2.64	**0.85**	1.51	0.94	1.69
	N=50 T=10	**0.95**	45.78	0.94	45.91	**0.94**	4.69	0.95	4.80
	N=50 T=50	**0.17**	1.61	0.17	1.62	**0.17**	0.32	0.17	0.32
M2	N=10 T=10	46.82	**57.34**	48.10	58.74	46.54	**25.20**	47.92	28.49
	N=10 T=50	46.77	**2.48**	48.04	2.64	46.86	**1.51**	48.11	1.69
	N=50 T=10	47.76	**45.78**	48.00	45.91	47.71	**4.69**	47.98	4.80
	N=50 T=50	47.74	**1.61**	47.98	1.62	47.75	**0.32**	47.99	0.32
M3	N=10 T=10	20.31	79.86	**5.40**	58.74	24.37	56.13	**5.52**	28.49
	N=10 T=50	3.16	5.09	**0.92**	2.64	3.85	4.59	**0.94**	1.69
	N=50 T=10	13.41	72.01	**0.94**	45.91	17.73	39.64	**0.95**	4.80
	N=50 T=50	2.41	4.35	**0.17**	1.62	3.02	3.56	**0.17**	0.32
M4	N=10 T=10	36.30	79.86	48.10	**58.74**	36.09	56.13	47.92	**28.49**
	N=10 T=50	36.75	5.09	48.04	**2.64**	37.13	4.59	48.11	**1.69**
	N=50 T=10	37.51	72.01	48.00	**45.91**	37.25	39.64	47.98	**4.80**
	N=50 T=50	37.62	4.35	47.98	**1.62**	37.87	3.56	47.99	**0.32**

Note: Numbers in the main entries are 1000×MSEs of the estimates of β.

Table 5: Application I: Crime rates in North Carolina (N=90, T=7, k=17)

	AIC	BIC	BIC_2	CV	CV^*	CV^{**}
Model 1	-2.121	-2.001	-2.125	0.124	0.094	0.028
Model 2	-3.773	-3.025	-3.796	0.025	0.023	0.026
Model 3	-2.124	-1.962	-2.129	0.124	0.094	0.027
Model 4	-3.823	-3.032	-3.847	0.024	0.022	0.025
Selected model	M4	M4	M4	M4	M4	M4

Table 6: Application II: Cross-country saving rates (N=56, T=15, k=5)

	AIC	BIC	BIC_2	CV
Model 1	2.547	2.576	2.547	12.844
Model 2	2.505	2.843	2.498	12.459
Model 3	2.555	2.663	2.553	12.953
Model 4	2.512	2.929	2.504	12.584
Selected model	M2	M1	M2	M2

Table 7: Application III: Guns and crime in the U.S.

	Static models (N=51, T=23, k=9)						Dynamic models (N=51, T=22, k=10)			
Model	AIC	BIC	BIC_2	CV	CV^*	CV^{**}	AIC	BIC	BIC_2	CV
log (violent crime rate)										
M1	-1.6911	-1.6522	-1.6914	0.1860	0.0165	0.0073	-4.8520	-4.8072	-4.8524	0.0078
M2	-3.6072	-3.3523	-3.6094	0.0274	0.0080	0.0072	-4.8719	-4.6033	-4.8746	0.0077
M3	-1.7198	-1.5859	-1.7210	0.1816	0.0140	0.0061	-5.0845	-4.9457	-5.0859	0.0062
M4	-3.8653	-3.5154	-3.8684	0.0211	0.0063	0.0059	-5.1235	-4.7609	-5.1271	0.0060
Selected	M4	M4	M4	M4	M4	M4	M4	M3	M4	M4
log (murder rate)										
M1	-1.6202	-1.5813	-1.6205	0.1991	0.1234	0.0560	-2.8836	-2.8388	-2.8841	0.0561
M2	-2.9845	-2.7296	-2.9867	0.0510	0.0457	0.0452	-3.1044	-2.8358	-3.1071	0.0453
M3	-1.7012	-1.5673	-1.7024	0.1844	0.1144	0.0550	-2.9087	-2.7699	-2.9101	0.0548
M4	-3.1243	-2.7744	-3.1274	0.0443	0.0413	0.0421	-3.1913	-2.8287	-3.1950	0.0415
Selected	M4	M4	M4	M4	M4	M4	M4	M1	M4	M4
log (robbery rate)										
M1	-0.9853	-0.9464	-0.9856	0.3748	0.0375	0.0164	-4.0919	-4.0472	-4.0924	0.0168
M2	-3.0239	-2.7690	-3.0261	0.0490	0.0167	0.0156	-4.1352	-3.8666	-4.1379	0.0161
M3	-1.1079	-0.9740	-1.1091	0.3338	0.0305	0.0137	-4.2892	-4.1505	-4.2906	0.0138
M4	-3.2181	-2.8682	-3.2212	0.0403	0.0135	0.0130	-4.3454	-3.9828	-4.3491	0.0131
Selected	M4	M4	M4	M4	M4	M4	M4	M3	M4	M4