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

March 12, 2018

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.

*We gratefully thank the Co-editor Oliver Linton, an associate editor and three anonymous referees for their many constructive comments. We thank the participants at the 2017 China Meeting of the Econometric Society (Wuhan) and the workshops at Chinese University of Hong Kong and Sun Yat-sen University for their helpful comments. We thank Silin Huang for her excellent research assistance. Lu gratefully acknowledges the funding support from the Hong Kong University of Science and Technology under grant numbers SBI15BM21 and SBI16BM28. 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, time effects, both, or neither in linear panel data models. Specifically, we consider the following four models:

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

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 despite the fact they can be either fixed effects or random effects for our purpose. 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 individual or time 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]

There are two main motivations for model selection in the above panel setup. First, we usually can achieve a small mean squared error (MSE) for the estimators of parameters of interest based on the true model, as shown in our simulation results in Section 3 (see Tables 4A and 4C). Therefore it is desirable to use the true model for point estimation and inference. In our simulations, we also show that the MSEs based on the selected model are usually smaller than those based on a single fixed model by ignoring the true underlying data generating process (DGP). Second, sometimes we may be interested in knowing whether the individual/time effects are present, as

[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.

these effects represent the unobservable heterogeneity and may have economic meaning. For example, in the wage equation where y_{it} is the hourly wage and x_{it} contains variables such as education and working experience, among others, the individual effects may be thought of as individual's unobservable ability. We may be interested in knowing whether the "ability" variable enters the wage equation.

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. In all information-criterion-based methods, there is an implicit tuning parameter (e.g., a Bayesian information criterion (BIC) specifies the penalty term to be proportional to $\ln(NT)/(NT)$, which works as a tuning parameter). There, to show the consistency of model selection, we often have the flexibility of choosing alternative tuning parameters. For the procedure based on hypothesis testing as discussed below, we need to choose the sequence of testing and the nominal level, which are difficulty to choose in practice. Second, we assume that the cross-section dimension (N) and time dimension (T) pass to infinity simultaneously but allow the relative rate between N and T to be arbitrary. For example, T can be much slower than N such as $T \asymp \ln(N)$. Although our method requires a relatively large T for the asymptotic analysis, it can be applied to the 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. Sixth, in our simulations, we show that our jackknife outperforms other competing methods, such as AIC and BIC in the absence of serial correlation in the error terms. In the presence of strong serial correlation, only our modified jackknife works well and other methods, such as jackknife, AIC and BIC, all break down.

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

[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), Shao (1993), Burman, Chow and Nolan (1994), Racine (2000), Hansen and Racine (2012), and Lu and Su (2015)).

[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.

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. 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.

In this paper, we only focus on the consistency of model selection and do not address the issue of post-selection inference. As is well known in the literature, usually post-selection inference is not uniformly valid (see, e.g., Leeb and Pötscher (2005)). This is a general and challenging question in the model selection literature. Despite its importance, it is beyond the scope of this

[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.

paper to provide a thorough theory on uniform inference.

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. Section 5 concludes. The proofs of the main results are relegated to Appendix A and some additional materials are included in the online supplement.

Notation. For an $m \times n$ real matrix A, we denote its transpose as A' and its Frobenius norm as $\|A\|$ ($\equiv [\mathrm{tr}(AA')]^{1/2}$) where $\equiv$ means "is defined as". Let $P_A \equiv A\,(A'A)^{-1}\,A'$ 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:

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

where $\underline{\alpha} = (\alpha_1, ..., \alpha_{N-1})'$, $\underline{\lambda} = (\lambda_1, ..., \lambda_{T-1})'$, $\beta^{(1)} = \beta$, $\beta^{(2)} = \left(\beta', \underline{\alpha}'\right)'$, $\beta^{(3)} = \left(\beta', \underline{\lambda}'\right)'$, 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

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

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 $\{(y_{it}, x_{it}^{(m)}) : 1 \le i \le N,\ 1 \le t \le T\}$ 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

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

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.

Remark 1. In dynamic panel models with individual effects (Models 2 and 4), it is well known that bias-correction is needed for the estimation and inference of parameters unless $T/N \to \infty$ as $(N, T) \to \infty$. There is a large literature on this issue; see, e.g., Arellano and Hahn (2007) and

Fernandez-Val and Weidner (2018) for a review. Here our purpose is to select the correct model consistently. We show that our jackknife method can select the true model consistently without the need for bias-correction. Our simulations also suggest that bias-correction for dynamic panels may or may not help with the determination of the true model.[5] Of course, we generally need to implement the bias-correction for estimation and inference after the model selection.

Remark 2. Here we focus on the determination of whether the individual effects, time effects, both, or neither should enter the model from the out-of-sample predictive power of these effects. Even though we treat either effects as fixed parameters to be estimated and allow them to be correlated with the regressors in x_{it}, they can be either fixed effects or random effects for subsequent estimation and inference. Our jackknife method can only tell whether either the individual effects or time effects are present or not, but cannot tell whether they are random or fixed effects. Note that even in a random effects model, we have the issue of which effects should be included in the model. For example, in an experimental setting where the key regressor is randomized, we still need to consider whether we should include individual or time effects (or both) for the efficiency consideration. Our method provides a practical solution. In a setting where it is unclear whether we should use random or fixed effects, we may implement a two-step procedure. In the first step, we apply our method to determine whether the individual or time effects should be included in the model. In the second step, we apply the Hausman-Wu type test (see, e.g., Hausman (1978) and Hausman and Taylor (1981)) to determine whether the effects are "random" or "fixed".

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) $E(u_{it}) = 0$, $\max_{1\leq i\leq N, 1\leq t\leq T} E(u_{it}^4) \leq C$, and $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it}^2 \xrightarrow{P} \bar{\sigma}_u^2 > 0$.

(ii) $\max_{1\leq i\leq N, 1\leq t\leq T} E \|x_{it}\|^4 \leq C$.

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

(iv) 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 $D_{\alpha\lambda}$.

(v) $\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} \alpha_i = o_P(1)$ 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.

[5] The simulation details are available upon request.

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)-(ii) impose weak moment conditions on $\{u_{it}\}$ and $\{x_{it}\}$, which are frequently assumed in the literature. The fourth-order moment conditions on $\{u_{it}\}$ and $\{x_{it}\}$ imply that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|x_{it}u_{it}\|^2 = O_P(1)$ and $\max_{1\le i\le N, 1\le t\le T}\|x_{it}\| = O_P((NT)^{1/4})$ by Markov and Jensen inequalities and the union bound. Assumption A.1(iii) is also weak and commonly imposed 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(iv) 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 correspondingly and then we can redefine $x_{it}^{(m)}$ for $m = 2, 3, 4$ with such regressors removed. For example, if x_{it} contains a time-invariant regressor other than the constant term, say, z_i, then z_i will be omitted from Models 2 and 4 in the estimation procedure, but still kept in Models 1 and 3. The omission of z_i in Models 2 and 4 will not cause the endogenous problem, as its effect will be captured by the individual effects in Models 2 and 4, which are allowed to be correlated with the other regressors in x_{it}. So the asymptotic analysis below will continue to hold. Assumption A.1(v) 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 a 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 $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E(u_{it}u_{is})$ 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 $\alpha_i(\tau) = \tau^{-\gamma_i}$ with $\min_{1\leq i\leq N}\gamma_i > (2+\delta)/\delta$; 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. In the online supplement, we demonstrate that the primitive conditions to ensure Assumption A.2(iii)-(iv) are: (i) $\max_{1\leq i\leq N} E\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it}^{*}u_{is}\right\|^2 \leq C$, (ii) $\max_{1\leq t\leq T} E\left\|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{it}^{*}u_{jt}\right\|^2 \leq C$, and (iii) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E(u_{it}u_{js})| \leq C$, where $x_{it}^{*} = x_{it} - E(x_{it})$. Analogous conditions are frequently assumed in the panel data literature to control weak serial and cross-section dependence; see, e.g., Bai and Ng (2002). It is worth mentioning that $\bar{\sigma}_{u1}^2 = \bar{\sigma}_u^2$ if there is no serial correlation among $\{u_{it}, t \geq 1\}$, and $\bar{\sigma}_{u2}^2 = \bar{\sigma}_u^2$ if there is no cross-sectional correlation among $\{u_{it}, i \geq 1\}$. 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 or misspecified 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, Models 1 and 3 are underfitted and misspecified, respectively. 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. In short, Assumption A.3 rules out asymptotic multicollinearity between the individual/time effects and the regressors.

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*

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

Remark 3. The proof of Theorem 2.1 is given in the 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 $P\left(CV(1) < CV(m)\right) \to 1$ for $m = 2, 3, 4$ by showing that

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

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

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

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

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

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

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

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

Remark 4. 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 $T = O(\ln N)$ such that our method may be applied to micro panels when T is typically small in comparison with N.

Remark 5. To see when the above additional condition can be met in Theorem 2.1, consider 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 and is independently and identically distributed (i.i.d.) along the cross-section dimension. 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 $\mathrm{Corr}(u_{it}, u_{jt}) = \rho^{|i-j|}$ 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 $\rho < \frac{1}{3}$. We can also consider a stationary m-dependent process for $\{u_{it}, t \geq 1\}$ with mean zero and variance σ_u^2 for each i (assuming i.i.d. in the cross-section dimension). In this case, we can show that

$$\bar{\sigma}_{u1}^2 = \lim_{T\to\infty}\left(\sigma_u^2 + \frac{2(T-1)}{T}\sum_{j=1}^{m}\mathrm{Cov}\left(u_{i1}, u_{i,j+1}\right)\right) = \left[1 + 2\sum_{j=1}^{m}\mathrm{Corr}\left(u_{i1}, u_{i,j+1}\right)\right]\sigma_u^2.$$

So the condition $\bar{\sigma}_{u1}^2 < 2\bar{\sigma}_u^2$ is satisfied if $\sum_{j=1}^{m} \text{Corr}(u_{i1}, u_{i,j+1}) < \frac{1}{2}$. Again, this means that we cannot have too large time-series correlation. Note that this condition is always satisfied for an invertible MA(1) process: $u_{it} = e_{it} + \theta e_{i,t-1}$ where $|\theta| < 1$ and $\{e_{it}, t \geq 1\}$ is a white noise with mean 0 and variance σ_e^2, as in this case,

$$\bar{\sigma}_{u1}^2 = \lim_{T\to\infty} \left(1 + \frac{T-1}{T} \frac{2\theta}{1+\theta^2}\right) \sigma_u^2 < 2\sigma_u^2,$$

where $\sigma_u^2 = (1+\theta^2)\sigma_e^2$.

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. Note that if a generic ARMA process is invertible, it can be written as an AR(∞) process and well approximated by an AR(p) process for sufficiently large p. For this reason, we assume that the error process $\{u_{it}, t \geq 1\}$ can be approximated by an AR(p) process:

$$\begin{aligned} u_{it} &= \sum_{j=1}^{\infty} \rho_j u_{i,t-j} + e_{it} = \left(\sum_{j=1}^{p} \rho_j u_{i,t-j}\right) + \left(\sum_{j=p+1}^{\infty} \rho_j u_{i,t-j} + e_{it}\right) \\ &= \boldsymbol{\rho}' \underline{u}_{i,t-1} + v_{it}, \end{aligned} \tag{2.10}$$

where $i = 1, ..., N$, $t = p+1, ..., T$, $\boldsymbol{\rho} = (\rho_1, ..., \rho_p)'$ is a vector of unknown parameters, $\underline{u}_{i,t-1} = (u_{i,t-1}, ..., u_{i,t-p})'$, $v_{it} = v_{it,p} + e_{it}$, e_{it} is an innovation term, and $v_{it,p} = \sum_{j=p+1}^{\infty} \rho_j u_{i,t-j}$ signifies the approximation error. If $\{u_{it}, t \geq 1\}$ is an autoregressive process of order p or less, then $v_{it,p} = 0$ and $v_{it} = e_{it}$.

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 $T_p = T - p$. Let $\underline{u}_{i,t-1} = (u_{i,t-1}, ..., u_{i,t-p})'$ and $\Gamma_p = \frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \underline{u}_{i,t-1} \underline{u}_{i,t-1}'$.

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

Assumption A.4. (i) $\sum_{j=1}^{\infty} \rho_j z^j \neq 0$ for any complex number z with $|z| \leq 1$, $\sum_{j=1}^{\infty} |\rho_j| < \infty$, $p^{3/2}(N^{-1} + T^{-1}) = o(1)$, and $\lambda_{\min}(\Gamma_p)$ is bounded away from zero in probability as $(N, T) \to \infty$.

(ii) $E(v_{it}) = 0$, $\max_{1 \leq i \leq N, p+1 \leq t \leq T} E(v_{it}^4) \leq C$, 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) $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \zeta_{it} v_{it} = O_P((NT)^{-1/2})$ for $\zeta_{it} = 1$ and $\breve{x}_{it}$, and $\left\| \frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \underline{u}_{i,t-1} v_{it} \right\| = O_P((NT/p)^{-1/2})$.

(iv) $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it} \alpha_i = o_P(1)$ and $\frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it} [\Phi(L) \lambda_t] = o_P(1)$ 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) $\frac{1}{N} \sum_{i=1}^{N} \bar{x}_{i\cdot} \bar{v}_{i\cdot} = O_P(T^{-1} + (NT)^{-1/2})$.

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

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

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

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

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

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

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

(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

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

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)-(iv) parallels Assumption A.1(i), (iii) 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). In the online supplement (Section C.2), we verify Assumptions A.4 and A.5 under a set of sufficient primitive conditions on $\{(x_{it},e_{it},\alpha_i,\lambda_t)\}$.

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*

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

Remark 6. Theorem 2.2 indicates that the modified jackknife criterion function helps us to select the correct model w.p.a.1 as $(N,T)\rightarrow\infty$ 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. More generally, if $\{u_{it},t\geq 1\}$ is an AR(∞) process, in the online supplement, we show that when $p\rightarrow\infty$ under certain rate condition, the approximation error $v_{it,p}(=\sum_{j=p+1}^{\infty}\rho_j u_{i,t-j})$ is asymptotically negligible so that $\bar{\sigma}^2_{v1}=\bar{\sigma}^2_{v2}=\bar{\sigma}^2_v$ is always satisfied.

Remark 7. 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 8. 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

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

where $\tilde{x}_{it} = \left(1, \mathring{x}_{it}',\ y_{i,t-1},\ \dots,\ y_{i,t-p},\ \mathring{x}_{i,t-1}',\ \dots,\ \mathring{x}_{it-p}'\right)'$, $\tilde{\beta}$ is the new vector of regression coefficients, $\tilde{\alpha}_i = \left(1 - \rho_1 - \dots - \rho_p\right)\alpha_i$ and $\tilde{\lambda}_t = \left(\lambda_t - \rho_1 \lambda_{t-1} - \dots - \rho_p \lambda_{t-p}\right)$. 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 9. As mentioned above, we regard our AR(p) model as an approximation for the error process $\{u_{it}, t \geq 1\}$ that does 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 $\{u_{it}, t \geq 1\}$. Despite this fact, we need to choose the value of p. In practice there are several approaches. First, we may use a "rule of thumb" and let p increase with T, e.g., $p = \lfloor T^{1/4} \rfloor$, where $\lfloor T^{1/4} \rfloor$ is the nearest integer less than or equal to $T^{1/4}$. Alternatively, we can follow Lee, Okui, and Shintani (2018) by setting $p_{\max} = \lfloor T^{1/4} \rfloor$ and consider a general-to-specific testing procedure based on t-statistic until we reject the null. Third, we may apply the information criteria, such as AIC and BIC, to the residuals obtained from Model 4 ($\hat{u}_{it}^{(4)}$) to determine p. For the implementation, see, e.g., Stock and Watson (2012, Section 14.5). In general, BIC gives a consistent estimator of p, and AIC tends to choose a relatively large p. See Section D in the online supplement for more details.

Remark 10. As a referee points out, the standard jackknife (cross-validation) method is originally designed for i.i.d. observations. For dependent time series data, various modifications have been proposed in the literature. For example, Burman, Chow and Nolan (1994) consider a h-block cross-validation function by removing the tth observation and the h observations on its either side to estimate the regression parameter, which simplifies to the usual leave-one-out cross-validation function when $h = 0$. Racine (2000) finds that the h-block cross-validation is not consistent in general and proposes to combine Shao's (1993) solution of v-blocking on independent data with Burman, Chow and Nolan's (1994) h-blocking on dependent data to yield a hv-block cross-validation for improved model-selection. Note that the h-block cross-validation requires the selection of one tuning parameter (h) while the hv-block cross-validation requires the choices of two tuning parameters (h and v). The minimum sample size (T in our notation) in Racine's (2000) simulations is 100 in order for his method to work reasonably well. But we usually do not have so many time series observations in the panel setup. If T is only 5, 10, or at most 50 as in our simulations, we do not know how these alternative methods work and whether it is possible to justify their consistency in determining whether to include the individual or time effects into a panel data model. At the minimum, our modified jackknife method offers an easy-to-implement alternative solution to handle serial correlation in the error terms that only demands the choice of a single tuning parameter (p) in practice.

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

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

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

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

We choose the model by minimizing the above three ICs.[6]

[6] 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.

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 8 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 : $(N,T) = (10,5)\,,\ (50,5)\,,\ (10,10)\,,\ (50,10)\,,\ (10,50)$ and $(50,50)\,.$ The number of replications is 1000.

3.2 Static panel models

We consider the following static fixed-effect 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 $\frac{3}{4}$.[7] 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 best. For example, when $N = 10$ and $T = 10$, our CV can choose the correct model with a probability close to 95%. When T is small ($T = 5$) and N is large ($N = 50$), the correct rate of our CV is above 90%. Even when both T and N are small ($T = 5, N = 10$), our CV can achieve a reasonable correct rate above 70%. 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 when T is relatively large ($T = 10$ or $T = 50$). 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 4%. BIC$_2$ outperforms BIC, but still underperforms CV and AIC.

Next, we consider $\rho = \frac{1}{4}$, i.e., there is weak serial correlation in the error term. When T is relatively large ($T = 10$ or $T = 50$), 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**. When T is small, CV** can perform poorly. 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). When T is small ($T = 5$), CV can even outperform CV*

[7] Here the error terms are homoskedastic. We have also considered the DGPs with heteroskedasticy and find that performances are similar. The details are available upon request.

and CV**. 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 56% and 44%, 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 in large samples, 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. However, in general, for CV* and CV** to work well, a relatively large T is required.

When the serial 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 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%. Also, when T or N is small, CV* and CV** can perform poorly. This suggests that when serial correlation is high, a large sample is required.

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 the four different models and the model selected by our CV when $\rho = 0$.[8] 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 as large as that based on Model 1. When T is relatively large, the MSEs based on our selected models are almost the same as those based on the true models. When T is small, our model selection can also achieve MSE reduction, compared with say, the largest model, Model 4. Table 4B reports the performance of post-selection inference by presenting the empirical coverage and length of the 95% confidence intervals (CI). We find that for this DGP, the empirical coverage and length based on our selected model are similar to those based on the true models, especially when T is relatively large.

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 in large samples and CV* generally outperforms CV**. Also it is noted that a relatively large T is required for CV* and

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

CV** to work well in the presence of high correlation.

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% and 80%, 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 4C shows the MSEs of estimator of β based on the four models and the selected model by CV when $\beta = \frac{3}{4}$.[9] 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. The MSEs based on our selected models are close to those based on the true models when T is large. When T is small, in general, our model selection can also achieve a smaller MSE than a single fixed model. Table 4D shows that the empirical coverage and length of the 95% CI based on our selected model are comparable to those based on the true models, especially when T is large.

[9] 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:

$$\text{DGP 3.1: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2 x_{it,j} + u_{it},$$
$$\text{DGP 3.2: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2 x_{it,j} + \alpha_i + u_{it},$$
$$\text{DGP 3.3: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2 x_{it,j} + \lambda_t + u_{it},$$
$$\text{DGP 3.4: } y_{it} = 1 + \beta y_{i,t-1} + \sum_{j=1}^{5} 0.2 x_{it,j} + \alpha_i + \lambda_t + u_{it},$$

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^{**}, where the number of lags p used in CV^* and CV^{**} is 1. All these methods determine that Model 4 (i.e., including both individual and time fixed effects) is the correct model. We also apply AIC and BIC to determine the number of lags p, both of which choose $p = 0$. Table 5 also reports the estimates and 95% CIs for the coefficient on the probability of arrest.[10] We consider both the non-clustered and clustered standard errors (SEs) where the clustered SEs are robust to the serial correlation in the error terms. Based on the selected Model 4, the point estimate is around -0.355 and the coefficient is significant at the 5% level.

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. Table 6 also present the estimation and inference results for the coefficient on the per capita GDP growth rate.[11] Given the dynamic specification, we report both the non-bias corrected and bias-corrected results. The bias-correction is based on the half panel jackknife method as proposed in Dhaene and Jochmans (2015). Based on the selected Model 2, the bias-corrected estimate is 0.178 with the 95% CI of [0.074, 0.281].

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

[10] The results for other coefficients are available upon request.
[11] The results for other coefficients are available upon request.

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 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). The number of lags p used in CV* and CV** is 1. We also apply AIC and BIC to determine the number of lags p. Both AIC and BIC determine the same value of p, and choose $p = 1$, 0 and 1 for the three dependent variables (the violent crime rate, the murder rate and the robbery rate), respectively. In this application, the coefficient on the "shall issue" is often the parameter of interest. Table 7 also reports the estimation and inference results. We find that the effect of the "shall issue" is not significant at the 5% level based on the selected Model 4. For example, when the dependent variable is the logarithm of the violent crime rate, the estimate is -0.028 with the 95% CI of [-0.106, 0.050] using the clustered SEs. In this application, whether to include individual fixed effects makes a difference. If we do not include individual effects, the effects of the "shall issue" are in general negative and significant at the 5% level. However, after including individual fixed effects, the significance is gone.

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 (2017) for a review). Therefore, it is even more imperative to select an appropriate specification of fixed effects in multidimensional panels. Second, given the fact that there is no natural ordering along the cross-section dimension in general, it is not easy to extend our jackknife method as in Section 2.3 to allow for strong or moderate degree of cross-section dependence in the standard two-way or one-way panel. If cross-section dependence is a concern, one can follow Bai (2009) and consider the determination of individual effects, time effects, and interactive fixed effects (IFEs) in the following model

$$y_{it} = x_{it}'\beta + \alpha_i + \lambda_t + \gamma_i' f_t + u_{it},$$

where f_t is an $R \times 1$ vector of factors and γ_i is an $R \times 1$ vector of factor loadings. The above equation models the cross-section dependence explicitly. Conceptually, we can apply the jackknife idea to the above models to select the number of factors and to determine the presence of α_i and/or λ_t simultaneously. The major difficulty lies in the fact that after deleting one observation, the panel data becomes unbalanced, which is not easy to deal with due to the presence of unobservable factors (see, e.g., Bai, Liao and Yang (2015)). When the regressors also share the factor structure as in Pesaran (2006), 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. Alternatively, we could model cross-sectional dependence using certain metric of economic distance, as in Conley (1999). We shall explore these topics in our future research.

Appendix

A Proofs of the main results

Let $\delta_{NT} = N^{-1} + T^{-1}$. To prove Theorem 2.1, we need the following six lemmas whose proofs can be found in the online supplement.

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

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

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

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*

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

where $X_D^* = \left(X'M_DX\right)^{-1}$ and $B_\ell = X'D_\ell\left(D_\ell'D_\ell\right)^{-1}$ for $\ell = 1, 2$.

Lemma A.3 *Suppose that Assumption A.1(iii) holds. Then*

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

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

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

Lemma A.4 *Suppose that Assumptions A.1(iii) 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) $\frac{1}{NT}X'D_{\alpha\lambda}\left(D_{\alpha\lambda}'D_{\alpha\lambda}\right)^{-1}D_{\alpha\lambda}'U = \frac{1}{N}\sum_{i=1}^N \bar{x}_{i\cdot}\bar{u}_{i\cdot} + \frac{1}{T}\sum_{t=1}^T \bar{x}_{t\cdot}\bar{u}_{t\cdot} - 2\bar{x}_{\cdot\cdot}\bar{u}_{\cdot\cdot} = O_P(\delta_{NT})$.

Lemma A.5 *Let $\eta_{it}^{(m)} = x_{it}^{(m)\prime}\left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U$ and $J_{mNT} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\eta_{it}^{(m)})^2 = \frac{1}{NT}U'X^{(m)} \times \left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U$ for $m = 1, 2, 3, 4$. Suppose that Assumptions A.1(iii)-(iv) 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) $J_{4NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(\delta_{NT}^2\right)$, $J_{4NT} - J_{2NT} = \frac{1}{T}\sum_{t=1}^T \bar{u}_{\cdot t}^2 + O_P\left(N^{-2} + (NT)^{-1}\right)$, *and* $J_{4NT} - J_{3NT} = \frac{1}{N}\sum_{i=1}^N \bar{u}_{i\cdot}^2 + O_P\left(T^{-2} + (NT)^{-1}\right)$.

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

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

(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* $\max_{i,t} h_{it}^{(2)}=O_P(T^{-1}+(NT)^{-1/2})$,
(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* $\max_{i,t} h_{it}^{(3)}=O_P(N^{-1}+(NT)^{-1/2})$,
(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* $\max_{i,t} h_{it}^{(4)}=O_P(\delta_{NT})$,
(v) $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(x_{it}-B_\alpha d_{\alpha,it})' X_{D_\alpha}^{*}(x_{it}-B_\alpha d_{\alpha,it})u_{it}^2=O_P((NT)^{-1})$,
(vi) $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(x_{it}-B_\lambda d_{\lambda,it})' X_{D_\lambda}^{*}(x_{it}-B_\lambda d_{\lambda,it})u_{it}^2=O_P((NT)^{-1})$,
(vii) $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(x_{it}-B_{\alpha\lambda} d_{\alpha\lambda,it})' X_{D_{\alpha\lambda}}^{*}(x_{it}-B_{\alpha\lambda} d_{\alpha\lambda,it})u_{it}^2=O_P((NT)^{-1})$.

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

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

where $h_{it}^{(m)}=x_{it}^{(m)\prime}\left(X^{(m)\prime}X^{(m)}\right)^{-1}x_{it}^{(m)}$. 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 $c_{it,m}=(1-h_{it}^{(m)})^{-1}$ and $c_{it,lm}=c_{it,l}c_{it,m}$. By Lemma A.6, for $l,m=1,2,3,4$ we have

$$\max_{i,t}h_{it}^{(m)}=O_P(\delta_{mNT}),\ \ \max_{i,t}|c_{it,m}-1|=O_P(\delta_{mNT})\ \text{ and }\ \max_{i,t}|c_{it,lm}-1|=O_P(\delta_{lNT}+\delta_{mNT}),\tag{A.2}$$

where $\delta_{1NT}=(NT)^{-1/2}$, $\delta_{2NT}=T^{-1}+(NT)^{-1/2}$, $\delta_{3NT}=N^{-1}+(NT)^{-1/2}$, and $\delta_{4NT}=\delta_{NT}$.

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

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

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 $\beta^{(m)}$, $m=2,3,4$, are all zero. This, in conjunction with (A.1), implies that for $m=1,2,3,4$,

$$\begin{aligned}
x_{it}^{(m)\prime}(\hat\beta_{it}^{(m)}-\beta^{(m)}) &= x_{it}^{(m)\prime}\Bigg[(\hat\beta^{(m)}-\beta^{(m)})-\frac{1}{1-h_{it}^{(m)}}\left(X^{(m)\prime}X^{(m)}\right)^{-1}x_{it}^{(m)}u_{it}\\
&\quad+\frac{1}{1-h_{it}^{(m)}}\left(X^{(m)\prime}X^{(m)}\right)^{-1}x_{it}^{(m)}x_{it}^{(m)\prime}B_U^{(m)}\Bigg]\\
&= x_{it}^{(m)\prime}B_U^{(m)}-\frac{h_{it}^{(m)}}{1-h_{it}^{(m)}}u_{it}+\frac{h_{it}^{(m)}}{1-h_{it}^{(m)}}x_{it}^{(m)\prime}B_U^{(m)}\\
&= -\frac{h_{it}^{(m)}}{1-h_{it}^{(m)}}u_{it}+\frac{1}{1-h_{it}^{(m)}}x_{it}^{(m)\prime}B_U^{(m)},
\end{aligned}$$

and

$$y_{it}-\hat y_{it}^{(m)}=u_{it}-x_{it}^{(m)\prime}\left(\hat\beta_{it}^{(m)}-\beta^{(m)}\right)=c_{it,m}\left(u_{it}-x_{it}^{(m)\prime}B_U^{(m)}\right).\tag{A.3}$$

It follows that $CV_{1,m} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{y}_{it}^{(m)}\right)^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} c_{it,m}^2 \left(u_{it} - x_{it}^{(m)\prime} B_U^{(m)}\right)^2$. We first study $CV_{1,2} - CV_{1,1}$. We make the following decomposition:

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

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) and (v) we can readily show that

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

This result, in conjunction with (A.2) 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.2)

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

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

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

By Lemma A.5(ii), $A_{2,1} = \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 + O_P((NT)^{-1}+T^{-2})$, where the first term is $O_P(T^{-1})$. This result, in conjunction with (A.2) 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

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

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

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

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

$$T\left[CV_{1,2}-CV_{1,1}\right] = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2 - \frac{T}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 + o_P(1) \xrightarrow{P} 2\bar{\sigma}_u^2 - \bar{\sigma}_{u1}^2, \tag{A.4}$$

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

$$N\left[CV_{1,3}-CV_{1,1}\right] = 2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2 - \frac{N}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2 + o_P(1) \xrightarrow{P} 2\bar{\sigma}_u^2 - \bar{\sigma}_{u2}^2, \tag{A.5}$$

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

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

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

$$
\begin{aligned}
A_4 &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(h_{it}^{(4)}-h_{it}^{(1)})u_{it}^2+o_P\left(N^{-1}+T^{-1}\right) \\
&= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{N-1}{NT}+\frac{T-1}{NT}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)' X_{D_{\alpha\lambda}}^{*}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)\right]u_{it}^2+o_P\left(N^{-1}+T^{-1}\right) \\
&= \left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2+O_P((NT)^{-1})+o_P\left(N^{-1}+T^{-1}\right) \\
&= \left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2+o_P\left(N^{-1}+T^{-1}\right), \\
A_5 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^{(4)\prime}B_U^{(4)}\right)^2+o_P\left(N^{-1}+T^{-1}\right)=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2+o_P\left(N^{-1}+T^{-1}\right), \\
A_6 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}x_{it}^{(2)\prime}B_U^{(2)}+o_P\left(N^{-1}+T^{-1}\right)=\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2+\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2+o_P\left(N^{-1}+T^{-1}\right).
\end{aligned}
$$

It follows that

$$
\begin{aligned}
(N\wedge T)\left[CV_{1,4}-CV_{1,1}\right] &= (N\wedge T)\left[\left(T^{-1}+N^{-1}\right)\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2-\frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2-\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^2\right]+o_P(1) \\
&\xrightarrow{P} 2(1+c)\,\bar{\sigma}_u^2-\left(\bar{\sigma}_{u1}^2+c\bar{\sigma}_{u2}^2\right)1\{c_1\geq 1\}-\left(c\bar{\sigma}_{u1}^2+\bar{\sigma}_{u2}^2\right)1\{c_1<1\}, \quad \text{(A.6)}
\end{aligned}
$$

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.4)-(A.6) 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, 3 and 4 are underfitted, misspecified and overfitted, respectively, 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.3), we can show that

$$
y_{it}-\hat{y}_{it}^{(1)}=u_{\alpha,it}-x_{it}'\left(\hat{\beta}_{it}^{(1)}-\beta^{(1)}\right)=c_{it,1}\left(u_{\alpha,it}-x_{it}'B_{U_\alpha}^{(1)}\right). \quad \text{(A.7)}
$$

where $B_{U_\alpha}^{(m)}=\left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U_\alpha$ for $m=1,2,3,4$. Then

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

It is easy to show that by Assumptions A.1 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.2) 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

$$CV_{2,1} - CV_{2,2} \xrightarrow{P} c_{\alpha,X} > 0. \tag{A.8}$$

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

$$y_{it} - \hat{y}_{it}^{(3)} = u_{\alpha,it} - x_{it}^{(3)\prime}\left(\hat{\beta}_{it}^{(3)} - \beta^{(3)}\right) = c_{it,3}\left(u_{\alpha,it} - x_{it}^{(3)\prime}B_{U_\alpha}^{(3)}\right), \tag{A.9}$$

and

$$\begin{aligned} CV_{2,3} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{\alpha,it} - x_{it}^{(3)\prime}B_{U_\alpha}^{(3)}\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(c_{it,3}^2 - 1\right)\left(u_{\alpha,it} - x_{it}^{(3)\prime}B_{U_\alpha}^{(3)}\right)^2 \\ &\equiv A_9 + A_{10}, \text{ say.} \end{aligned}$$

By Assumptions A.1(i), A.1(iii) and A.3(i), Lemmas A.4-A.5, and (A.2), we can readily show that

$$\begin{aligned} A_9 &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_i - x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_\alpha\underline{\alpha} + u_{it} - x_{it}^{(3)\prime}B_U^{(3)}\right)^2 \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\alpha_i - x_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_\alpha\underline{\alpha}\right)^2 + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^2 + o_P(1) \xrightarrow{P} c_{\alpha,X_\lambda} + \bar{\sigma}_u^2, \end{aligned}$$

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

$$CV_{2,3} - CV_{2,2} \xrightarrow{P} c_{\alpha,X_\lambda} > 0. \tag{A.10}$$

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

$$y_{it} - \hat{y}_{it}^{(4)} = u_{it} - x_{it}^{(4)\prime}\left(\hat{\beta}_{it}^{(4)} - \beta^{(4)}\right) = c_{it,4}\left(u_{it} - x_{it}^{(4)\prime}B_U^{(4)}\right), \tag{A.11}$$

we have

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

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.2), we can readily show that

$$
\begin{aligned}
A_{11} &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(h_{it}^{(4)}-h_{it}^{(2)}\right)u_{it}^{2}+o_P\left(N^{-1}\right)\\
&= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{T-1}{NT}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)'X_{D_{\alpha\lambda}}^{*}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)-\left(x_{it}-B_{\alpha}d_{\alpha,it}\right)'X_{D_{\alpha}}^{*}\left(x_{it}-B_{\alpha}d_{\alpha,it}\right)\right]u_{it}^{2}\\
&\quad +o_P\left(N^{-1}\right)\\
&= N^{-1}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}+o_P\left(N^{-1}\right),\\
A_{12} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(x_{it}^{(4)\prime}B_U^{(4)}\right)^2-\left(x_{it}^{(2)\prime}B_U^{(2)}\right)^2\right]+o_P\left(N^{-1}\right)=\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_P\left(N^{-1}\right),\text{ and}\\
A_{13} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[u_{it}x_{it}^{(4)\prime}B_U^{(4)}-u_{it}x_{it}^{(2)\prime}B_U^{(4)}\right]+o_P\left(N^{-1}\right)=\frac{1}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_P\left(N^{-1}\right).
\end{aligned}
$$

It follows that

$$
N\left[CV_{2,4}-CV_{2,2}\right]=\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^{2}-\frac{N}{T}\sum_{t=1}^{T}\bar{u}_{\cdot t}^{2}+o_P(1)\xrightarrow{P}2\bar{\sigma}_u^2-\bar{\sigma}_{u2}^2, \tag{A.12}
$$

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

By (A.8), (A.10), and (A.12), 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 $\underline{\lambda}=(\lambda_1,...,\lambda_{T-1})'$. Let $B_{U_\lambda}^{(m)}=\left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U_\lambda$ and $B_{U_{\alpha\lambda}}^{(m)}=\left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U_{\alpha\lambda}$ for $m=1,2,3,4$. Following the steps to obtain (A.3), now we can show that

$$
y_{it}-\hat{y}_{it}^{(1)}=u_{\alpha\lambda,it}-x_{it}'(\hat{\beta}_{it}^{(1)}-\beta^{(1)})=c_{it,1}\left(u_{\alpha\lambda,it}-x_{it}'B_{U_{\alpha\lambda}}^{(1)}\right). \tag{A.13}
$$

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

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

Similarly, we have by Assumptions A.1 and A.3(iii)

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

and

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

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 three new lemmas. The proofs of these lemmas can be found in the online supplement. 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 $\mathbf{u}_i = (u_{i,p+1}, ..., u_{i,T})'$, $\mathbf{U} = (\mathbf{u}_1', ..., \mathbf{u}_N')'$, $\mathbf{z}_i = (\underline{\ddot{u}}_{i,p}, ..., \underline{\ddot{u}}_{i,T-1})'$ 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 $\bar{u}_{\cdot\cdot}$. Let $\breve{x}_{it}^{(m)} = x_{it}^{(m)} - \underline{x}_{i,t-1}^{(m)}\boldsymbol{\rho}$ where $\underline{x}_{i,t-1}^{(m)} = (x_{i,t-1}^{(m)}, ..., x_{it-p}^{(m)})$ for $m = 1,2,3,4$.

Lemma A.7 *Suppose Assumptions A.1, A.2 and A.4 hold. Then* $\|\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}\| = O_P(p\delta_{NT})$.

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 2.2. Noting that $(y_{it}-\hat{\boldsymbol{\rho}}'\underline{y}_{i,t-1})-(\hat{y}_{it}^{(m)}-\hat{\boldsymbol{\rho}}'\hat{\underline{y}}_{i,t-1}^{(m)})=(y_{it}-\boldsymbol{\rho}'\underline{y}_{i,t-1})-(\hat{y}_{it}^{(m)}-\boldsymbol{\rho}'\hat{\underline{y}}_{i,t-1}^{(m)})$ $+(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho})'(\hat{\underline{y}}_{i,t-1}^{(m)}-\underline{y}_{i,t-1})$, we have

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

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.3)

$$\begin{aligned}
(y_{it}-\boldsymbol{\rho}'\underline{y}_{i,t-1})-(\hat{y}_{it}^{(m)}-\boldsymbol{\rho}'\hat{\underline{y}}_{i,t-1}^{(m)}) &= c_{it,m}[u_{it}-x_{it}^{(m)\prime}B_U^{(m)}]-\sum_{j=1}^{p}\rho_j c_{i,t-j,m}[u_{i,t-j}-x_{i,t-j}^{(m)\prime}B_U^{(m)}] \\
&= c_{it,m}[v_{it}-\breve{x}_{it}^{(m)\prime}B_U^{(m)}]+\sum_{j=1}^{p}\rho_j\varkappa_{it,m,j}[u_{i,t-j}-x_{i,t-j}^{(m)\prime}B_U^{(m)}], \qquad \text{(A.14)}
\end{aligned}$$

where $c_{it,m}=(1-h_{it}^{(m)})^{-1}$, and $\varkappa_{it,m,j}=c_{it,m}-c_{i,t-j,m}$ for $m=1,2,3,4$ and $j=1,...,p$. By Lemma A.6, we have

$$\max_{i,t}|\varkappa_{it,m,j}|=O_P(\delta_{mNT}) \text{ for } m=1,2,3,4 \text{ and } j=1,...,p. \qquad \text{(A.15)}$$

Note that

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

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

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

where we use the fact that $\bar{v}_{i\cdot}=\frac{1}{T_p}\sum_{t=p+1}^T v_{it}=\frac{1}{T_p}\sum_{t=p+1}^T\left(u_{it}-\underline{u}_{i,t-1}'\boldsymbol{\rho}\right)=\Phi(1)\bar{u}_{i\cdot}+O_P\left(T^{-1}\right)$. Similarly, using (A.15) and following the analysis of $CV_{1,2}-CV_{1,1}$, we can readily show that $T_p[CV_{1,2}^*(1,2)-CV_{1,1}^*(1,2)]=o_P(1)$ and $T_p\left[CV_{1,2}^*(1,3)-CV_{1,1}^*(1,3)\right]=o_P(1)$. It follows that $T_p\left[CV_{1,2}^*(1)-CV_{1,1}^*(1)\right]\xrightarrow{P}2\bar{\sigma}_v^2-\bar{\sigma}_{v1}^2$.

By (A.3) and (A.14),

$$\begin{aligned}
CV_{1,m}^*(2) &= \frac{(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho})'}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T(\underline{\hat{y}}_{i,t-1}^{(m)}-\underline{y}_{i,t-1})(\underline{\hat{y}}_{i,t-1}^{(m)}-\underline{y}_{i,t-1})'(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho})\\
&= \sum_{j_1=1}^p\sum_{j_2=1}^p(\hat{\rho}_{j_1}-\rho_{j_1})(\hat{\rho}_{j_2}-\rho_{j_2})D_{1,m}(1,j_1,j_2),\ \text{and}\\
CV_{1,m}^*(3) &= \frac{(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho})'}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T(\underline{\hat{y}}_{i,t-1}^{(m)}-\underline{y}_{i,t-1})\left[(y_{it}-\boldsymbol{\rho}'\underline{y}_{i,t-1})-(\hat{y}_{it}^{(m)}-\boldsymbol{\rho}'\underline{\hat{y}}_{i,t-1}^{(m)})\right]\\
&= \sum_{j_1=1}^p(\hat{\rho}_{j_1}-\rho_{j_1})\left\{D_{1,m}(2,j_1)+D_{1,m}(3,j_1)\right\},
\end{aligned}$$

where

$$\begin{aligned}
D_{1,m}(1,j_1,j_2) &= \frac{1}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T c_{i,t-j_1,m}c_{i,t-j_2,m}[u_{i,t-j_1}-x_{i,t-j_1}^{(m)\prime}B_U^{(m)}][u_{i,t-j_2}-x_{i,t-j_2}^{(m)\prime}B_U^{(m)}],\\
D_{1,m}(2,j_1) &= \frac{1}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T c_{i,t-j_1,m}c_{it,m}[u_{i,t-j_1}-x_{i,t-j_1}^{(m)\prime}B_U^{(m)}]\left[v_{it}-\breve{x}_{it}^{(m)\prime}B_U^{(m)}\right],\\
D_{1,m}(3,j_1) &= \frac{1}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T\sum_{j=1}^p\rho_j c_{i,t-j_1,m}\varkappa_{it,m,j}[u_{i,t-j_1}-x_{i,t-j_1}^{(m)\prime}B_U^{(m)}][u_{i,t-j}-x_{i,t-j}^{(m)\prime}B_U^{(m)}].
\end{aligned}$$

As in the analysis of $CV_{1,2}-CV_{1,1}$, we can readily show that $D_{1,2}(1,j_1,j_2)-D_{1,1}(1,j_1,j_2)=O_P\left(T^{-1}\right)$ and $D_{1,m}(\ell,j_1)=O_P\left((NT)^{-1}\right)$ for $\ell=2,3$ uniformly in $j_1,j_2=1,...,p$. Then by Lemma A.7

$$\begin{aligned}
T_p\left[CV_{1,2}^*(2)-CV_{1,1}^*(2)\right] &= p\left\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\right\|^2O_P(1)=O_P(p^3\delta_{NT}^2)=o_P(1),\ \text{and}\\
T_p\left[CV_{1,2}^*(3)-CV_{1,1}^*(3)\right] &= \sqrt{p}\left\|\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}\right\|O_P(1)=O_P(p^{3/2}\delta_{NT})=o_P(1).
\end{aligned}$$

In sum, we have

$$
\begin{aligned}
T_p\left[CV_{1,2}^*-CV_{1,1}^*\right] &= \left(\frac{2}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T v_{it}^2-\frac{T_p}{N}\sum_{i=1}^N\bar{v}_{i\cdot}^2\right)+\frac{T_p}{N}\sum_{i=1}^N\left[\bar{v}_{i\cdot}-\Phi(1)\bar{u}_{i\cdot}\right]^2+o_P(1)\\
&\overset{P}{\to} 2\bar{\sigma}_v^2-\bar{\sigma}_{v1}^2. \qquad \text{(A.16)}
\end{aligned}
$$

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

$$
\begin{aligned}
T_p\left[CV_{1,3}^*-CV_{1,1}^*\right] &= \left(\frac{2}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T v_{it}^2-\frac{N}{T_p}\sum_{t=p+1}^T\bar{v}_{\cdot t}^2\right)+\frac{N}{T_p}\sum_{t=p+1}^T\left[\bar{v}_{\cdot t}-\Phi(L)\bar{u}_{\cdot t}\right]^2+o_P(1)\\
&\overset{P}{\to} 2\bar{\sigma}_v^2-\bar{\sigma}_{v2}^2, \qquad \text{(A.17)}
\end{aligned}
$$

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),

$$
\begin{aligned}
(N\wedge T_p)\left[CV_{1,4}-CV_{1,1}\right] &= (N\wedge T_p)\left\{\left(T_p^{-1}+N^{-1}\right)\frac{2}{NT_p}\sum_{i=1}^N\sum_{t=p+1}^T v_{it}^2-\frac{1}{N}\sum_{i=1}^N\bar{v}_{i\cdot}^2-\frac{1}{T_p}\sum_{t=p+1}^T\bar{v}_{\cdot t}^2\right\}+o_P(1)\\
&\overset{P}{\to} 2(1+c)\bar{\sigma}_v^2-\left(\bar{\sigma}_{v1}^2+c\bar{\sigma}_{v2}^2\right)1\{c_1\geq 1\}-\left(c\bar{\sigma}_{v1}^2+\bar{\sigma}_{v2}^2\right)1\{c_1<1\}, \qquad \text{(A.18)}
\end{aligned}
$$

where $c=\lim_{(N,T)\to\infty}\left(\frac{N}{T_p}\wedge\frac{T_p}{N}\right)$ and $c_1=\lim_{(N,T)\to\infty}\frac{N}{T_p}$. Combining (A.16)-(A.18) yields $P\left(CV_{1,1}^*<CV_{1,m}^*\right)$ $\to 1$ for $m=2,3,4$ provided $\max\left(\bar{\sigma}_{v1}^2,\bar{\sigma}_{v2}^2\right)<2\bar{\sigma}_v^2$.

Case 2: Model 2 is the true model. In this case, Model 1 is underfitted, Model 3 is misspecified, 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.7), we can show that

$$
\begin{aligned}
(y_{it}-\hat{y}_{it}^{(1)})-\boldsymbol{\rho}'(\underline{y}_{i,t-1}-\underline{\hat{y}}_{i,t-1}^{(1)}) &= c_{it,1}[u_{\alpha,it}-x_{it}'B_{U_\alpha}^{(1)}]-\sum_{j=1}^p\rho_j c_{i,t-j,1}[u_{\alpha,it-j}-x_{i,t-j}'B_{U_\alpha}^{(1)}]\\
&= c_{it,1}[\Phi(1)\alpha_i+v_{it}-\breve{x}_{it}'B_{U_\alpha}^{(1)}]+\sum_{j=1}^p\rho_j\varkappa_{it,1,j}[u_{\alpha,it-j}-x_{i,t-j}'B_{U_\alpha}^{(1)}]. \qquad \text{(A.19)}
\end{aligned}
$$

where $B_{U_\alpha}^{(m)}=\left(X^{(m)\prime}X^{(m)}\right)^{-1}X^{(m)\prime}U_\alpha$ for $m=1,2,3,4$. Then

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

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

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

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

$$
\begin{aligned}
(y_{it} - \hat{y}_{it}^{(2)}) - \boldsymbol{\rho}'(\underline{y}_{i,t-1} - \underline{\hat{y}}_{i,t-1}^{(2)}) &= c_{it,2}\left[u_{it} - x_{it}^{(2)\prime}B_U^{(2)}\right] - \sum_{j=1}^{p}\rho_j c_{i,t-j,2}\left[u_{i,t-j} - x_{i,t-j}^{(2)\prime}B_U^{(2)}\right] \\
&= c_{it,2}\left[v_{it} - \breve{x}_{it}^{(2)\prime}B_U^{(2)}\right] + \sum_{j=1}^{p}\rho_j \varkappa_{it,2,j}\left[u_{i,t-j} - x_{i,t-j}^{(2)\prime}B_U^{(2)}\right],
\end{aligned}
$$

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

$$
CV_{2,1}^* - CV_{2,2}^* \xrightarrow{P} c_{\alpha,X}^* > 0. \tag{A.20}
$$

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

$$
\begin{aligned}
(y_{it} - \hat{y}_{it}^{(3)}) - \boldsymbol{\rho}'(\underline{y}_{i,t-1} - \underline{\hat{y}}_{i,t-1}^{(3)}) &= c_{it,3}[u_{\alpha,it} - x_{it}^{(3)\prime}B_{U_\alpha}^{(3)}] - \sum_{j=1}^{p}\rho_j c_{i,t-j,3}[u_{\alpha,it-j} - x_{i,t-1}^{(3)\prime}B_{U_\alpha}^{(3)}] \\
&= c_{it,3}[\Phi(1)\alpha_i + v_{it} - \breve{x}_{it}^{(3)\prime}B_{U_\alpha}^{(3)}] + \sum_{j=1}^{p}\rho_j \varkappa_{it,3,j}[u_{\alpha,it-j} - x_{i,t-j}^{(3)\prime}B_{U_\alpha}^{(3)}], \quad \text{(A.21)}
\end{aligned}
$$

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

$$
\begin{aligned}
CV_{2,3}^* &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[(y_{it} - \hat{y}_{it}^{(3)}) - \boldsymbol{\rho}'(\underline{y}_{i,t-1} - \underline{\hat{y}}_{i,t-1}^{(3)})\right]^2 \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\Phi(1)\alpha_i + v_{it} - \breve{x}_{it}^{(3)\prime}B_{U_\alpha}^{(3)}\right]^2 + o_P(1) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\Phi(1)\alpha_i - \breve{x}_{it}^{(3)\prime}\left(X^{(3)\prime}X^{(3)}\right)^{-1}X^{(3)\prime}D_\alpha\underline{\alpha}\right]^2 + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}^2 + o_P(1) \\
&\xrightarrow{P} c_{\alpha,X_\lambda}^* + \bar{\sigma}_v^2,
\end{aligned}
$$

It follows that

$$
CV_{2,3}^* - CV_{2,2}^* \xrightarrow{P} c_{\alpha,X_\lambda}^* > 0. \tag{A.22}
$$

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

$$
\begin{aligned}
(y_{it} - \hat{y}_{it}^{(4)}) - \boldsymbol{\rho}'(\underline{y}_{i,t-1} - \underline{\hat{y}}_{i,t-1}^{(4)}) &= c_{it,4}[u_{it} - x_{it}^{(4)\prime}B_U^{(4)}] - \sum_{j=1}^{p}\rho_j c_{i,t-j,4}[u_{i,t-j} - x_{i,t-j}^{(4)\prime}B_U^{(4)}] \\
&= c_{it,4}[v_{it} - \breve{x}_{it}^{(4)\prime}B_U^{(4)}] + \sum_{j=1}^{p}\rho_j \varkappa_{it,4,j}[u_{i,t-j} - x_{i,t-j}^{(4)\prime}B_U^{(4)}], \quad \text{(A.23)}
\end{aligned}
$$

we have

$$
\begin{aligned}
& CV^*_{2,4} - CV^*_{2,2} \\
= \; & \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[c^2_{it,4}\left(v_{it}-\breve{x}^{(4)\prime}_{it}B^{(4)}_U\right)^2 - c^2_{it,2}\left(v_{it}-\breve{x}^{(2)\prime}_{it}B^{(2)}_U\right)^2\right] \\
& +\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left[\left(\sum_{j=1}^{p}\rho_j\varkappa_{it,4,j}[u_{i,t-j}-x^{(4)\prime}_{i,t-j}B^{(4)}_U]\right)^2 - \left(\sum_{j=1}^{p}\rho_j\varkappa_{it,2,j}[u_{i,t-j}-x^{(2)\prime}_{i,t-j}B^{(2)}_U]\right)^2\right] \\
& +\frac{2}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\sum_{j=1}^{p}\rho_j\left[c_{it,4}\varkappa_{it,4,j}(v_{it}-\breve{x}^{(4)\prime}_{it}B^{(4)}_U)(u_{i,t-j}-x^{(4)\prime}_{i,t-j}B^{(4)}_U)\right. \\
& \left.-c_{it,2}\varkappa_{it,2,j}(v_{it}-\breve{x}^{(2)\prime}_{it}B^{(2)}_U)(u_{i,t-j}-x^{(2)\prime}_{i,t-j}B^{(2)}_U)\right] \\
\equiv \; & D_{2,4}(1)+D_{2,4}(2)+2D_{2,4}(3)\,, \text{ say.}
\end{aligned}
$$

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

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

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.2) and (A.15), we can readily show that

$$
\begin{aligned}
D_{2,4}(1,1) &= N^{-1}\frac{2}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v^2_{it}+o_P\left(N^{-1}\right), \\
D_{2,4}(1,2) &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}[(\breve{x}^{(4)\prime}_{it}B^{(4)}_U)^2-(\breve{x}^{(2)\prime}_{it}B^{(2)}_U)^2]+o_P\left(N^{-1}\right)=\frac{1}{T_p}\sum_{t=p+1}^{T}\left[\Phi(L)\bar{u}_{\cdot t}\right]^2+o_P\left(N^{-1}\right), \\
D_{2,4}(1,3) &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}[\breve{x}^{(4)\prime}_{it}B^{(4)}_U-\breve{x}^{(2)\prime}_{it}B^{(2)}_U]+o_P\left(N^{-1}\right)=\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}\Phi(L)\bar{u}_{\cdot t}+o_P\left(N^{-1}\right).
\end{aligned}
$$

It follows that $N\cdot D_{2,4}(1)=\frac{2}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v^2_{it}-\frac{N}{T_p}\sum_{t=p+1}^{T}v^2_{\cdot t}+\frac{N}{T_p}\sum_{t=p+1}^{T}\left[v_{\cdot t}-\Phi(L)\bar{u}_{\cdot t}\right]^2+o_P(1)$. 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)

$$
N\left[CV^*_{2,4}-CV^*_{2,2}\right]=\frac{2}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v^2_{it}-\frac{N}{T_p}\sum_{t=p+1}^{T}v^2_{\cdot t}+o_P(1)\xrightarrow{P}2\bar{\sigma}^2_v-\bar{\sigma}^2_{v2}. \tag{A.24}
$$

By (A.20), (A.22), and (A.24), we have $P\left(CV^*_{2,2}<CV^*_{2,m}\right)\to 1$ as $(N,T)\to\infty$ for $m=1,3,4$ provided $\bar{\sigma}^2_{v2}<2\bar{\sigma}^2_v$.

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.14), now we can show that

$$\begin{aligned} & (y_{it} - \hat{y}_{it}^{(1)}) - \boldsymbol{\rho}'(\underline{y}_{i,t-1} - \underline{\hat{y}}_{i,t-1}^{(1)}) \\ = \; & c_{it,1}[u_{\alpha\lambda,it} - x_{it}' B_{U_{\alpha\lambda}}^{(1)}] - \sum_{j=1}^{p} \rho_j c_{i,t-j,1}[u_{\alpha\lambda,i,t-j} - x_{i,t-j}' B_{U_{\alpha\lambda}}^{(1)}] \\ = \; & c_{it,1}[\Phi(1)\alpha_i + \Phi(L)\lambda_t + v_{it} - \breve{x}_{it}' B_{U_{\alpha\lambda}}^{(1)}] + \sum_{j=1}^{p} \rho_j \varkappa_{it,1,j}[u_{\alpha\lambda,i,t-j} - x_{i,t-j}' B_{U_{\alpha\lambda}}^{(1)}], \end{aligned} \tag{A.25}$$

where $B_{U_{\alpha\lambda}}^{(m)} = \left(X^{(m)\prime} X^{(m)}\right)^{-1} X^{(m)\prime} U_{\alpha\lambda}$ for $m = 1,2,3,4$. As in Case 2, we can show that by Assumptions A.4(ii)-(iv) and A.6(iii),

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

Similarly, we have

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

and $CV_{4,4}^* = \frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} v_{it}^2 + o_P(1) \xrightarrow{P} \bar{\sigma}_v^2$. 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.

Arellano, M., and J. Hahn (2007). Understanding bias in nonlinear panel models: some recent developments. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics. In: Ninth World Congress, vol. III. Cambridge Press, pp.381–409.

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.

Bai, J., Y. Liao, and J. Yang (2015). Unbalanced panel data models with interactive effects. In B.H. Baltagi (ed.), The Oxford Handbook of Panel Data, pp.149-170.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models, Econometrica 70, 191-221.

Balazsi, L., L. Matyas, and T. Wansbeek (2017). Fixed effects models. In L. Matyas (eds), The Econometrics of Multi-dimensional Panels. 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.

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.

Burman, P., E. Chow, and D. Nolan (1994). A cross-validatory method for dependent data. Biometrika 84, 351-358.

Conley, T. (1999). GMM estimation with cross sectional dependence, Journal of Econometrics 92, 1-45.

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.

Fernández-Val, I. and M. Weidner (2018). Fixed effect estimation of large T panel data models, The Annual Review of Economics, forthcoming.

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.

Hausman, J. (1978). Specification tests in econometrics, Econometrica 46, 1251-1272.

Hausman, J. and W. Taylor (1981). Panel data and unobservable individual effects, Econometrica 49, 1377-1398.

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.

Lee, Y. J., R. Okui, and M. Shintani (2018). Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes, Journal of Econometrics, forthcoming.

Leeb, H. and B. Pötscher (2005). Model selection and inference: Facts and fiction, Econometric Theory 21, 21-59.

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.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: *hv*-block cross-validation, Journal of econometrics 99, 39-61.

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.

Shao, J. (1993). Linear model selection by cross-validation. Journal of the American statistical Association 88, 486-494.

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 M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.80**	.06	.11	.03	.04	**.74**	.02	.20	.05	.01	**.82**	.13	.06	.05	.04	**.86**
	(50,5)	**.90**	0	.10	0	0	**.84**	0	.16	.01	0	**.99**	0	0	0	0	**1**
	(10,10)	**.90**	.06	.04	.01	.01	**.91**	0	.08	.01	0	**.90**	.09	0	.01	.01	**.98**
	(50,10)	**.96**	0	.04	0	0	**.93**	0	.07	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.97**	.03	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.99**	0	.01	0	.49	**.46**	.04	.01	.23	0	**.77**	0	.74	.02	.06	**.18**
	(50,5)	**1**	0	0	0	.51	**0**	.49	0	.03	0	**.97**	0	.52	0	.48	**0**
	(10,10)	**1**	0	0	0	.14	**.86**	0	0	.20	0	**.80**	0	.74	.03	.01	**.22**
	(50,10)	**1**	0	0	0	.41	**.39**	.20	0	0	0	**1**	0	.54	0	.41	**.05**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.35	.28	**.37**	0	.40	.56	0	**.04**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC_2	(10,5)	**.44**	.20	.17	.20	0	**.59**	.01	.41	.02	.02	**.56**	.40	0	.04	.01	**.95**
	(50,5)	**.85**	.01	.14	.01	0	**.76**	0	.24	0	0	**.98**	.02	0	0	0	**1**
	(10,10)	**.65**	.13	.16	.06	0	**.74**	0	.26	0	0	**.76**	.24	0	0	0	**1**
	(50,10)	**.93**	0	.07	0	0	**.90**	0	.11	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.94**	.06	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.87**	.03	.10	0	.08	**.80**	.03	.08	.07	.01	**.90**	.03	.15	.07	.08	**.70**
	(50,5)	**.90**	0	.10	0	0	**.90**	.01	.09	.01	0	**1**	0	0	0	.02	**.98**
	(10,10)	**.93**	.04	.03	0	.01	**.96**	0	.03	.01	0	**.95**	.04	.01	.02	.01	**.96**
	(50,10)	**.96**	0	.04	0	0	**.97**	0	.04	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.97**	.03	0	0	0	**1**	0	0	0	0	**.98**	.02	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV^*	(10,5)	**.61**	.26	.09	.04	.03	**.83**	.02	.12	.07	.04	**.65**	.24	.09	.11	.04	**.77**
	(50,5)	**.79**	.12	.08	.01	0	**.89**	0	.11	0	0	**.87**	.12	0	.01	0	**.99**
	(10,10)	**.81**	.14	.04	.01	.01	**.94**	0	.05	.01	.01	**.86**	.13	.01	.02	.01	**.96**
	(50,10)	**.94**	.01	.05	0	0	**.94**	0	.06	0	0	**.99**	.01	0	0	0	**1**
	(10,50)	**.96**	.04	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV^{**}	(10,5)	**.79**	.08	.12	.02	.35	**.46**	.12	.07	.15	.02	**.75**	.08	.17	.08	.41	**.34**
	(50,5)	**.89**	0	.11	0	.16	**.49**	.30	.06	.01	0	**.99**	0	.02	.01	.52	**.46**
	(10,10)	**.88**	.08	.04	.01	.04	**.90**	.01	.04	.02	0	**.90**	.07	.02	.03	.12	**.83**
	(50,10)	**.95**	0	.05	0	0	**.95**	0	.05	0	0	**1**	0	0	0	.01	**1**
	(10,50)	**.97**	.03	0	0	0	**1**	0	0	0	0	**.97**	.04	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
	(10,5)	**.53**	.31	.06	.10	.02	**.78**	.01	.20	.03	.02	**.51**	.43	.02	.05	.02	**.91**
	(50,5)	**.69**	.20	.06	.05	0	**.82**	0	.18	.01	0	**.73**	.26	0	0	0	**1**
AIC	(10,10)	**.61**	.32	.03	.04	.01	**.89**	0	.10	.01	0	**.58**	.42	0	.01	.01	**.98**
	(50,10)	**.76**	.19	.03	.02	0	**.92**	0	.09	0	0	**.77**	.24	0	0	0	**1**
	(10,50)	**.72**	.28	0	0	0	**1**	0	0	0	0	**.63**	.37	0	0	0	**1**
	(50,50)	**.86**	.15	0	0	0	**1**	0	0	0	0	**.83**	.17	0	0	0	**1**
	(10,5)	**.98**	.02	.01	0	.36	**.61**	.02	.01	.25	.01	**.72**	.03	.61	.04	.05	**.30**
	(50,5)	**1**	0	0	0	.53	**.01**	.46	0	.03	0	**.97**	0	.55	0	.45	**0**
BIC	(10,10)	**1**	0	0	0	.11	**.89**	0	0	.24	0	**.75**	.01	.65	.05	.01	**.29**
	(50,10)	**1**	0	0	0	.31	**.59**	.10	0	0	0	**1**	0	.51	0	.33	**.16**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.37	.39	**.24**	0	.38	.60	0	**.02**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	(10,5)	**.20**	.40	.07	.33	0	**.58**	0	.41	.01	.02	**.23**	.74	0	.03	0	**.96**
	(50,5)	**.40**	.41	.05	.14	0	**.76**	0	.24	0	0	**.42**	.58	0	0	0	**1**
BIC_2	(10,10)	**.34**	.43	.08	.16	0	**.74**	0	.26	0	0	**.37**	.63	0	0	0	**1**
	(50,10)	**.59**	.33	.04	.04	0	**.89**	0	.11	0	0	**.59**	.41	0	0	0	**1**
	(10,50)	**.65**	.35	0	0	0	**1**	0	.01	0	0	**.54**	.46	0	0	0	**1**
	(50,50)	**.88**	.12	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
	(10,5)	**.67**	.24	.06	.03	.04	**.86**	.02	.09	.05	.02	**.69**	.24	.09	.06	.05	**.80**
	(50,5)	**.84**	.08	.06	.01	0	**.90**	0	.10	.01	0	**.90**	.09	0	0	0	**1**
CV	(10,10)	**.69**	.27	.03	.01	.01	**.95**	0	.04	.01	.01	**.71**	.27	0	.02	.01	**.97**
	(50,10)	**.84**	.13	.03	.01	0	**.95**	0	.05	0	0	**.87**	.13	0	0	0	**1**
	(10,50)	**.73**	.27	0	0	0	**1**	0	0	0	0	**.72**	.28	0	0	0	**1**
	(50,50)	**.87**	.13	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
	(10,5)	**.47**	.41	.06	.06	.04	**.82**	.02	.12	.04	.06	**.50**	.39	.07	.10	.04	**.80**
	(50,5)	**.51**	.40	.04	.04	0	**.90**	0	.11	0	0	**.56**	.43	0	0	0	**1**
CV^*	(10,10)	**.74**	.21	.04	.02	.01	**.94**	0	.05	.01	.01	**.77**	.21	.01	.02	.02	**.95**
	(50,10)	**.90**	.05	.05	0	0	**.95**	0	.05	0	0	**.95**	.05	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	(10,5)	**.75**	.12	.11	.03	.46	**.36**	.12	.06	.12	.02	**.75**	.11	.13	.06	.53	**.28**
	(50,5)	**.89**	.01	.11	0	.33	**.26**	.37	.03	.01	0	**.98**	.01	.01	0	.75	**.24**
CV^{**}	(10,10)	**.86**	.09	.04	.01	.18	**.74**	.04	.03	.02	.01	**.89**	.09	.02	.02	.35	**.61**
	(50,10)	**.95**	0	.05	0	.03	**.87**	.06	.05	0	0	**1**	0	0	0	.18	**.82**
	(10,50)	**.96**	.04	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
	(10,5)	**.41**	.42	.05	.13	.01	**.78**	.01	.20	.03	.03	**.38**	.56	.02	.05	.01	**.92**
	(50,5)	**.41**	.46	.03	.10	0	**.83**	0	.17	0	0	**.41**	.59	0	0	0	**1**
AIC	(10,10)	**.47**	.44	.03	.06	.01	**.89**	0	.11	.01	0	**.44**	.55	0	.01	.01	**.98**
	(50,10)	**.45**	.49	.02	.04	0	**.91**	0	.09	0	0	**.44**	.56	0	0	0	**1**
	(10,50)	**.57**	.43	0	0	0	**1**	0	0	0	0	**.48**	.52	0	0	0	**1**
	(50,50)	**.53**	.47	0	0	0	**1**	0	0	0	0	**.49**	.51	0	0	0	**1**
	(10,5)	**.95**	.04	.01	0	.31	**.66**	.02	.01	.27	.01	**.67**	.05	.55	.05	.04	**.36**
	(50,5)	**1**	0	0	0	.55	**.02**	.43	0	.04	0	**.97**	0	.57	0	.43	**0**
BIC	(10,10)	**.99**	.01	0	0	.10	**.90**	0	0	.27	.01	**.70**	.01	.61	.06	.01	**.32**
	(50,10)	**1**	0	0	0	.28	**.65**	.07	0	0	0	**1**	0	.49	0	.29	**.22**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.38	.45	**.17**	0	.37	.61	0	**.02**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	(10,5)	**.14**	.45	.05	.36	0	**.58**	0	.42	.01	.02	**.16**	.81	0	.03	0	**.97**
	(50,5)	**.15**	.63	.02	.20	0	**.76**	0	.24	0	0	**.15**	.85	0	0	0	**1**
BIC_2	(10,10)	**.24**	.52	.05	.20	0	**.74**	0	.25	0	0	**.25**	.75	0	0	0	**1**
	(50,10)	**.28**	.62	.02	.09	0	**.89**	0	.12	0	0	**.28**	.72	0	0	0	**1**
	(10,50)	**.47**	.52	.01	0	0	**.99**	0	.01	0	0	**.41**	.59	0	0	0	**1**
	(50,50)	**.60**	.40	0	0	0	**1**	0	0	0	0	**.56**	.44	0	0	0	**1**
	(10,5)	**.55**	.36	.05	.04	.03	**.87**	.01	.09	.05	.03	**.57**	.35	.07	.08	.04	**.82**
	(50,5)	**.64**	.29	.05	.03	0	**.90**	0	.11	.01	0	**.68**	.31	0	0	0	**1**
CV	(10,10)	**.54**	.42	.02	.02	.01	**.94**	0	.05	.01	.01	**.57**	.41	0	.02	.01	**.97**
	(50,10)	**.55**	.41	.02	.02	0	**.95**	0	.05	0	0	**.58**	.42	0	0	0	**1**
	(10,50)	**.58**	.42	0	0	0	**1**	0	0	0	0	**.58**	.43	0	0	0	**1**
	(50,50)	**.56**	.44	0	0	0	**1**	0	0	0	0	**.56**	.44	0	0	0	**1**
	(10,5)	**.42**	.47	.05	.07	.03	**.84**	.02	.12	.04	.06	**.44**	.47	.06	.09	.04	**.81**
	(50,5)	**.38**	.53	.03	.07	0	**.89**	0	.11	0	0	**.41**	.58	0	0	0	**1**
CV^*	(10,10)	**.69**	.25	.04	.02	.02	**.93**	0	.05	.01	0	**.73**	.26	.01	.02	.02	**.95**
	(50,10)	**.87**	.08	.04	.01	0	**.95**	0	.05	0	0	**.91**	.09	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.94**	.06	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	(10,5)	**.73**	.14	.11	.03	.48	**.35**	.12	.06	.11	.03	**.74**	.13	.11	.06	.55	**.28**
	(50,5)	**.88**	.01	.11	0	.38	**.23**	.37	.03	.01	0	**.98**	.01	.01	0	.79	**.19**
CV^{**}	(10,10)	**.84**	.11	.04	.01	.25	**.67**	.05	.03	.01	.01	**.88**	.10	.02	.01	.45	**.52**
	(50,10)	**.95**	0	.05	0	.09	**.75**	.13	.03	0	0	**1**	0	0	0	.34	**.66**
	(10,50)	**.96**	.04	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.18**	.62	.02	.18	.01	**.77**	0	.22	.02	.03	**.17**	.78	.01	.05	.01	**.94**
	(50,5)	**.02**	.79	0	.19	0	**.81**	0	.19	0	0	**.02**	.98	0	0	0	**1**
	(10,10)	**.21**	.68	.01	.10	.01	**.87**	0	.12	0	.01	**.18**	.81	0	.02	0	**.98**
	(50,10)	**.04**	.87	0	.09	0	**.91**	0	.09	0	0	**.03**	.97	0	0	0	**1**
	(10,50)	**.27**	.72	0	0	0	**1**	0	0	0	0	**.22**	.78	0	0	0	**1**
	(50,50)	**.04**	.96	0	0	0	**1**	0	0	0	0	**.03**	.97	0	0	0	**1**
BIC	(10,5)	**.78**	.21	0	.01	.21	**.77**	.01	.02	.27	.04	**.47**	.22	.42	.08	.03	**.47**
	(50,5)	**1**	0	0	0	.55	**.16**	.30	0	.05	0	**.95**	0	.62	0	.33	**.05**
	(10,10)	**.89**	.11	0	0	.09	**.91**	0	0	.36	.04	**.49**	.10	.51	.12	.01	**.36**
	(50,10)	**1**	0	0	0	.20	**.78**	.02	0	.01	0	**1**	0	.43	0	.16	**.42**
	(10,50)	**.98**	.02	0	0	0	**1**	0	0	.39	.58	**.03**	0	.37	.62	0	**.01**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC_2	(10,5)	**.05**	.55	.02	.39	0	**.59**	0	.41	0	.02	**.05**	.93	0	.02	0	**.97**
	(50,5)	**0**	.75	0	.25	0	**.75**	0	.25	0	0	**0**	1	0	0	0	**1**
	(10,10)	**.08**	.67	.02	.23	0	**.74**	0	.25	0	.01	**.09**	.90	0	.01	0	**.99**
	(50,10)	**.01**	.87	0	.12	0	**.88**	0	.13	0	0	**.01**	.99	0	0	0	**1**
	(10,50)	**.22**	.77	0	.01	0	**.99**	0	.01	0	0	**.18**	.82	0	0	0	**1**
	(50,50)	**.06**	.94	0	0	0	**1**	0	0	0	0	**.04**	.96	0	0	0	**1**
CV	(10,5)	**.28**	.63	.03	.07	.02	**.89**	0	.09	.04	.06	**.30**	.61	.03	.08	.02	**.87**
	(50,5)	**.07**	.82	.01	.11	0	**.88**	0	.12	0	0	**.08**	.92	0	.01	0	**1**
	(10,10)	**.25**	.69	.01	.05	.01	**.93**	0	.06	.01	.02	**.28**	.69	0	.03	.01	**.97**
	(50,10)	**.06**	.88	0	.06	0	**.94**	0	.06	0	0	**.06**	.94	0	0	0	**1**
	(10,50)	**.28**	.72	0	0	0	**1**	0	0	0	0	**.29**	.71	0	0	0	**1**
	(50,50)	**.04**	.96	0	0	0	**1**	0	0	0	0	**.05**	.95	0	0	0	**1**
CV^*	(10,5)	**.28**	.61	.04	.08	.03	**.84**	.01	.12	.03	.06	**.31**	.61	.04	.09	.04	**.84**
	(50,5)	**.13**	.76	.01	.10	0	**.88**	0	.12	0	0	**.14**	.85	0	0	0	**1**
	(10,10)	**.58**	.36	.03	.03	.03	**.91**	0	.06	.01	.01	**.61**	.37	.01	.01	.03	**.95**
	(50,10)	**.69**	.26	.03	.01	0	**.95**	0	.05	0	0	**.74**	.27	0	0	0	**1**
	(10,50)	**.91**	.09	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV^{**}	(10,5)	**.67**	.20	.10	.04	.50	**.34**	.11	.05	.08	.04	**.70**	.19	.09	.05	.58	**.28**
	(50,5)	**.86**	.03	.11	0	.49	**.20**	.29	.02	.01	0	**.96**	.04	.01	0	.82	**.17**
	(10,10)	**.81**	.14	.04	.01	.42	**.51**	.05	.02	.01	.01	**.84**	.14	.01	.01	.58	**.40**
	(50,10)	**.95**	.01	.05	0	.31	**.44**	.23	.02	0	0	**.99**	.01	0	0	.69	**.31**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	.02	**.98**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
	(10,5)	**.02**	.77	0	.21	.01	**.78**	0	.21	0	.05	**.02**	.93	0	.05	0	**.94**
	(50,5)	**0**	.80	0	.20	0	**.80**	0	.20	0	0	**0**	1	0	0	0	**1**
AIC	(10,10)	**.01**	.84	0	.14	0	**.85**	0	.15	0	.03	**.01**	.96	0	.03	0	**.97**
	(50,10)	**0**	.87	0	.13	0	**.87**	0	.13	0	0	**0**	1	0	0	0	**1**
	(10,50)	**.02**	.95	0	.03	0	**.97**	0	.03	0	.01	**.01**	.98	0	.01	0	**.99**
	(50,50)	**0**	.98	0	.02	0	**.98**	0	.02	0	0	**0**	1	0	0	0	**1**
	(10,5)	**.20**	.77	0	.03	.05	**.91**	0	.04	.11	.16	**.10**	.63	.13	.18	.01	**.68**
	(50,5)	**.52**	.48	0	.01	.06	**.93**	.01	.01	.06	.01	**.44**	.50	.10	.02	.03	**.86**
BIC	(10,10)	**.24**	.75	0	.01	.04	**.95**	0	.01	.19	.36	**.08**	.38	.19	.38	.01	**.42**
	(50,10)	**.47**	.53	0	0	.01	**.99**	0	0	.02	0	**.42**	.56	.06	.01	0	**.94**
	(10,50)	**.47**	.53	0	0	0	**1**	0	0	.25	.76	**0**	0	.23	.77	0	**0**
	(50,50)	**.77**	.23	0	0	0	**1**	0	0	0	0	**.74**	.26	0	0	0	**1**
	(10,5)	**.01**	.61	0	.38	0	**.62**	0	.38	0	.02	**.01**	.97	0	.02	0	**.98**
	(50,5)	**0**	.75	0	.25	0	**.75**	0	.25	0	0	**0**	1	0	0	0	**1**
BIC_2	(10,10)	**0**	.74	0	.26	0	**.74**	0	.26	0	.01	**.01**	.98	0	.01	0	**.99**
	(50,10)	**0**	.84	0	.16	0	**.84**	0	.16	0	0	**0**	1	0	0	0	**1**
	(10,50)	**.01**	.94	0	.05	0	**.95**	0	.05	0	.01	**.01**	.99	0	.01	0	**1**
	(50,50)	**0**	.98	0	.02	0	**.98**	0	.02	0	0	**0**	1	0	0	0	**1**
	(10,5)	**.04**	.85	0	.11	.01	**.88**	0	.11	.01	.08	**.04**	.87	.01	.08	.01	**.91**
	(50,5)	**0**	.87	0	.13	0	**.87**	0	.13	0	0	**0**	1	0	0	0	**1**
CV	(10,10)	**.02**	.90	.01	.08	0	**.92**	0	.08	.01	.06	**.03**	.91	0	.06	0	**.94**
	(50,10)	**0**	.91	0	.09	0	**.91**	0	.09	0	0	**0**	1	0	0	0	**1**
	(10,50)	**.02**	.96	0	.02	0	**.98**	0	.02	0	.02	**.03**	.95	0	.02	0	**.98**
	(50,50)	**0**	.98	0	.02	0	**.98**	0	.02	0	0	**0**	1	0	0	0	**1**
	(10,5)	**.10**	.77	.01	.12	.02	**.84**	.01	.13	.01	.07	**.11**	.81	.02	.08	.03	**.88**
	(50,5)	**0**	.87	0	.12	0	**.88**	0	.12	0	0	**0**	1	0	0	0	**1**
CV^*	(10,10)	**.30**	.63	.02	.04	.04	**.90**	0	.07	.01	0	**.32**	.68	0	.01	.05	**.94**
	(50,10)	**.15**	.79	.01	.05	0	**.94**	0	.06	0	0	**.17**	.83	0	0	0	**1**
	(10,50)	**.71**	.29	0	0	.01	**.99**	0	0	0	0	**.68**	.32	0	0	0	**1**
	(50,50)	**.96**	.04	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	0	**1**
	(10,5)	**.56**	.31	.08	.06	.49	**.36**	.09	.06	.06	.05	**.61**	.29	.06	.05	.58	**.31**
	(50,5)	**.66**	.22	.09	.03	.52	**.30**	.15	.04	0	0	**.76**	.23	0	.01	.69	**.30**
CV^{**}	(10,10)	**.66**	.29	.03	.02	.53	**.41**	.04	.02	.01	0	**.73**	.26	.01	0	.64	**.36**
	(50,10)	**.87**	.09	.04	0	.65	**.23**	.11	.01	0	0	**.91**	.09	0	0	.82	**.19**
	(10,50)	**.92**	.08	0	0	.32	**.68**	0	0	0	0	**.93**	.07	0	0	.52	**.48**
	(50,50)	**1**	0	0	0	.16	**.84**	0	0	0	0	**1**	0	0	0	.51	**.49**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.66**	.19	.08	.08	.04	**.74**	0	.21	.05	.01	**.65**	.29	0	.05	.04	**.91**
	(50,5)	**.88**	.03	.08	.01	0	**.85**	0	.15	0	0	**.95**	.04	0	0	0	**1**
	(10,10)	**.85**	.10	.04	.02	.01	**.91**	0	.09	.01	0	**.85**	.15	0	.01	0	**.99**
	(50,10)	**.96**	0	.04	0	0	**.93**	0	.07	0	0	**.99**	.01	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.92**	.08	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.99**	0	0	0	.71	**.27**	.01	.01	.21	0	**.79**	0	.14	.04	.56	**.27**
	(50,5)	**1**	0	0	0	1	**0**	0	0	.03	0	**.97**	0	.02	0	.98	**0**
	(10,10)	**1**	0	0	0	.41	**.59**	0	0	.11	0	**.89**	0	.07	.05	.31	**.57**
	(50,10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	0	0	1	**0**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.08	0	**.92**	0	0	.08	0	**.92**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC_2	(10,5)	**.28**	.34	.10	.28	0	**.60**	0	.40	.01	.01	**.36**	.62	0	.02	0	**.98**
	(50,5)	**.71**	.15	.09	.05	0	**.78**	0	.22	0	0	**.79**	.21	0	0	0	**1**
	(10,10)	**.57**	.22	.12	.09	0	**.76**	0	.24	0	0	**.65**	.34	0	0	0	**1**
	(50,10)	**.92**	.02	.06	0	0	**.89**	0	.11	0	0	**.97**	.03	0	0	0	**1**
	(10,50)	**.93**	.07	0	0	0	**1**	0	0	0	0	**.89**	.12	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.79**	.12	.08	.02	.09	**.80**	.01	.10	.07	.01	**.81**	.12	.01	.07	.13	**.79**
	(50,5)	**.91**	.01	.08	0	0	**.92**	0	.08	.01	0	**.99**	.01	0	0	.01	**.99**
	(10,10)	**.89**	.08	.03	.01	.01	**.95**	0	.04	.01	0	**.91**	.08	0	.01	.01	**.98**
	(50,10)	**.96**	0	.04	0	0	**.96**	0	.04	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.96**	.05	0	0	0	**1**	0	0	0	0	**.96**	.05	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.56**	.27	.06	.11	.06	**.73**	.01	.21	.04	.02	**.54**	.40	.01	.04	.06	**.90**
	(50,5)	**.78**	.12	.07	.03	0	**.85**	0	.15	0	0	**.84**	.15	0	0	0	**1**
	(10,10)	**.80**	.14	.04	.02	.01	**.91**	0	.09	.01	0	**.78**	.22	0	.01	.01	**.99**
	(50,10)	**.94**	.02	.04	0	0	**.93**	0	.07	0	0	**.98**	.02	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.92**	.09	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.98**	.01	0	0	.82	**.17**	.01	.01	.21	0	**.78**	.02	.15	.03	.65	**.18**
	(50,5)	**1**	0	0	0	1	**0**	0	0	.02	0	**.98**	0	.02	0	.98	**0**
	(10,10)	**1**	0	0	0	.76	**.24**	0	0	.12	0	**.88**	0	.10	.02	.64	**.24**
	(50,10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	0	0	1	**0**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.09	0	**.92**	0	0	.07	0	**.93**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC_2	(10,5)	**.19**	.42	.07	.32	0	**.60**	0	.40	.01	.02	**.25**	.73	0	.02	.01	**.97**
	(50,5)	**.48**	.35	.07	.10	0	**.77**	0	.23	0	0	**.53**	.47	0	0	0	**1**
	(10,10)	**.51**	.28	.09	.12	0	**.76**	0	.24	0	0	**.55**	.45	0	0	0	**1**
	(50,10)	**.88**	.05	.06	.01	0	**.90**	0	.10	0	0	**.94**	.06	0	0	0	**1**
	(10,50)	**.92**	.08	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.70**	.20	.07	.04	.13	**.75**	.02	.09	.06	.02	**.73**	.20	.01	.07	.22	**.70**
	(50,5)	**.88**	.04	.08	0	.03	**.89**	.01	.07	0	0	**.97**	.03	0	0	.05	**.95**
	(10,10)	**.84**	.12	.03	.01	.01	**.95**	0	.04	.01	0	**.88**	.11	0	.01	.03	**.96**
	(50,10)	**.95**	.01	.04	0	0	**.96**	0	.04	0	0	**.99**	.01	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.37**	.45	.04	.15	.09	**.70**	.01	.20	.03	.03	**.34**	.60	.01	.05	.08	**.87**
	(50,5)	**.44**	.44	.05	.07	.01	**.84**	0	.15	0	0	**.45**	.54	0	0	.01	**.99**
	(10,10)	**.63**	.30	.03	.03	.05	**.87**	0	.08	.01	0	**.61**	.39	0	0	.05	**.95**
	(50,10)	**.83**	.12	.03	.02	0	**.93**	0	.08	0	0	**.85**	.16	0	0	0	**1**
	(10,50)	**.92**	.08	0	0	0	**1**	0	0	0	0	**.89**	.12	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.97**	.03	0	0	.87	**.12**	0	0	.20	.01	**.76**	.04	.15	.02	.71	**.12**
	(50,5)	**1**	0	0	0	1	**0**	0	0	.03	0	**.98**	0	.02	0	.98	**0**
	(10,10)	**1**	0	0	0	.95	**.05**	0	0	.12	0	**.87**	.01	.09	.01	.85	**.05**
	(50,10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	0	0	1	**0**
	(10,50)	**1**	0	0	0	.08	**.92**	0	0	.09	0	**.91**	0	.03	.05	.08	**.84**
	(50,50)	**1**	0	0	0	.91	**.10**	0	0	0	0	**1**	0	0	0	.90	**.10**
BIC_2	(10,5)	**.09**	.52	.03	.36	0	**.60**	0	.40	0	.02	**.10**	.88	0	.03	0	**.97**
	(50,5)	**.13**	.65	.02	.20	0	**.77**	0	.23	0	0	**.14**	.86	0	0	0	**1**
	(10,10)	**.33**	.45	.05	.17	.01	**.75**	0	.24	0	0	**.35**	.65	0	0	.01	**.99**
	(50,10)	**.66**	.26	.04	.04	0	**.90**	0	.10	0	0	**.68**	.33	0	0	0	**1**
	(10,50)	**.89**	.11	0	0	0	**1**	0	0	0	0	**.84**	.17	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.54**	.36	.05	.05	.20	**.69**	.03	.08	.05	.03	**.58**	.34	.02	.07	.29	**.63**
	(50,5)	**.70**	.21	.07	.02	.09	**.83**	.01	.07	0	0	**.78**	.21	0	0	.12	**.88**
	(10,10)	**.71**	.26	.02	.01	.08	**.88**	0	.04	.01	0	**.76**	.23	0	.01	.14	**.85**
	(50,10)	**.89**	.07	.03	.01	.01	**.95**	0	.04	0	0	**.92**	.08	0	0	.01	**.99**
	(10,50)	**.93**	.07	0	0	0	**1**	0	0	0	0	**.93**	.07	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.59**	.18	.11	.13	.09	**.62**	.04	.26	.05	.02	**.60**	.33	.04	.05	.11	**.80**
	(50,5)	**.87**	.03	.09	.01	.01	**.77**	.02	.20	0	0	**.96**	.04	0	0	.05	**.95**
	(10,10)	**.81**	.11	.05	.03	.01	**.88**	.01	.11	0	0	**.80**	.20	.01	.01	.02	**.96**
	(50,10)	**.95**	0	.04	0	0	**.91**	0	.09	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.94**	.06	0	0	0	**1**	0	0	0	0	**.91**	.09	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.99**	.01	.01	0	.78	**.18**	.03	.02	.26	0	**.72**	.02	.49	.02	.36	**.13**
	(50,5)	**1**	0	0	0	.96	**0**	.04	0	.03	0	**.97**	0	.09	0	.92	**0**
	(10,10)	**1**	0	0	0	.68	**.32**	0	0	.23	0	**.77**	0	.71	.01	.17	**.11**
	(50,10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	.02	0	.98	**0**
	(10,50)	**1**	0	0	0	0	**1**	0	0	.38	.28	**.33**	0	.79	.18	0	**.03**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC_2	(10,5)	**.26**	.29	.10	.36	.01	**.49**	.01	.48	.02	.02	**.32**	.65	0	.03	.02	**.95**
	(50,5)	**.72**	.11	.12	.06	0	**.71**	0	.29	0	0	**.81**	.19	0	0	0	**1**
	(10,10)	**.55**	.21	.12	.13	0	**.71**	0	.28	0	0	**.62**	.38	0	0	0	**1**
	(50,10)	**.91**	.01	.07	0	0	**.87**	0	.13	0	0	**.98**	.02	0	0	0	**1**
	(10,50)	**.92**	.08	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.79**	.10	.09	.02	.31	**.54**	.07	.08	.10	.01	**.79**	.10	.16	.07	.35	**.43**
	(50,5)	**.92**	0	.08	0	.10	**.73**	.10	.08	.01	0	**.99**	0	0	0	.27	**.73**
	(10,10)	**.91**	.06	.02	.01	.03	**.93**	.01	.03	.01	.01	**.92**	.06	.03	.02	.09	**.86**
	(50,10)	**.96**	0	.04	0	0	**.96**	0	.04	0	0	**1**	0	0	0	0	**1**
	(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N.T)																
AIC	(10.5)	**.50**	.25	.09	.16	.11	**.59**	.04	.26	.05	.03	**.50**	.43	.03	.05	.13	**.79**
	(50.5)	**.81**	.08	.08	.03	.04	**.71**	.03	.22	0	0	**.87**	.12	0	0	.07	**.92**
	(10.10)	**.77**	.14	.05	.04	.04	**.85**	.01	.11	0	0	**.76**	.24	.01	.01	.06	**.93**
	(50.10)	**.95**	.01	.04	0	0	**.90**	0	.10	0	0	**.99**	.01	0	0	0	**1**
	(10.50)	**.94**	.06	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
	(50.50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10.5)	**.97**	.02	.01	0	.83	**.14**	.02	.01	.27	.01	**.70**	.03	.42	.02	.44	**.12**
	(50.5)	**1**	0	0	0	.99	**0**	.01	0	.03	0	**.97**	0	.05	0	.95	**0**
	(10.10)	**1**	0	0	0	.87	**.13**	0	0	.24	0	**.76**	0	.63	.01	.29	**.07**
	(50.10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	.01	0	.99	**0**
	(10.50)	**1**	0	0	0	.01	**1**	0	0	.40	.27	**.33**	0	.92	.06	0	**.02**
	(50.50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	.01	**.99**
BIC_2	(10.5)	**.19**	.34	.07	.40	.02	**.48**	.01	.49	.01	.02	**.23**	.73	0	.03	.02	**.94**
	(50.5)	**.56**	.22	.09	.13	0	**.68**	0	.32	0	0	**.62**	.38	0	0	.01	**.99**
	(10.10)	**.50**	.26	.10	.14	0	**.71**	.01	.28	0	0	**.55**	.45	0	0	.01	**.99**
	(50.10)	**.89**	.03	.07	.01	0	**.86**	0	.14	0	0	**.95**	.05	0	0	0	**1**
	(10.50)	**.91**	.09	0	0	0	**1**	0	0	0	0	**.87**	.14	0	0	0	**1**
	(50.50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10.5)	**.75**	.13	.08	.04	.37	**.48**	.07	.08	.10	.02	**.75**	.13	.14	.07	.40	**.40**
	(50.5)	**.90**	.01	.08	0	.23	**.58**	.10	.09	.01	0	**.98**	.01	0	0	.39	**.61**
	(10.10)	**.87**	.10	.02	.01	.08	**.88**	.01	.03	.01	.01	**.88**	.10	.03	.02	.18	**.78**
	(50.10)	**.96**	0	.04	0	0	**.95**	0	.04	0	0	**1**	0	0	0	.01	**.99**
	(10.50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	0	**1**
	(50.50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

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

True M		Model 1				Model 2				Model 3				Model 4			
Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	(N,T)																
AIC	(10,5)	**.37**	.38	.06	.20	.14	**.56**	.04	.26	.04	.04	**.35**	.57	.02	.06	.14	**.78**
	(50,5)	**.59**	.25	.06	.10	.10	**.64**	.02	.24	0	0	**.61**	.39	0	0	.11	**.89**
	(10,10)	**.64**	.27	.03	.06	.13	**.76**	.01	.10	0	0	**.61**	.38	.01	.01	.14	**.84**
	(50,10)	**.88**	.07	.03	.02	.04	**.85**	0	.11	0	0	**.89**	.11	0	0	.05	**.95**
	(10,50)	**.93**	.08	0	0	0	**1**	0	0	0	0	**.88**	.12	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	(10,5)	**.95**	.04	.01	0	.86	**.11**	.01	.01	.27	.01	**.66**	.06	.35	.02	.53	**.11**
	(50,5)	**1**	0	0	0	1	**0**	0	0	.03	0	**.97**	0	.04	0	.96	**0**
	(10,10)	**1**	0	0	0	.96	**.04**	0	0	.26	0	**.74**	0	.48	.01	.49	**.03**
	(50,10)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	.01	0	.99	**0**
	(10,50)	**1**	0	0	0	.38	**.62**	0	0	.44	.24	**.32**	0	.98	.01	0	**.01**
	(50,50)	**1**	0	0	0	1	**0**	0	0	0	0	**1**	0	0	0	1	**0**
BIC_2	(10,5)	**.11**	.41	.05	.45	.02	**.48**	.01	.49	.01	.03	**.13**	.83	0	.04	.03	**.93**
	(50,5)	**.23**	.47	.04	.27	.01	**.64**	0	.35	0	0	**.24**	.76	0	0	.01	**.99**
	(10,10)	**.36**	.39	.07	.18	.02	**.69**	.02	.27	0	0	**.37**	.63	0	0	.04	**.96**
	(50,10)	**.75**	.16	.05	.04	.01	**.83**	0	.16	0	0	**.78**	.22	0	0	.01	**.99**
	(10,50)	**.89**	.11	0	0	0	**1**	0	0	0	0	**.84**	.17	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
CV	(10,5)	**.65**	.23	.07	.05	.41	**.45**	.06	.07	.09	.04	**.66**	.21	.10	.07	.44	**.39**
	(50,5)	**.82**	.08	.08	.02	.36	**.48**	.07	.09	0	0	**.89**	.10	0	0	.47	**.53**
	(10,10)	**.78**	.19	.02	.01	.24	**.71**	.01	.03	.01	.01	**.81**	.17	.03	.01	.37	**.59**
	(50,10)	**.93**	.03	.03	0	.08	**.85**	.01	.06	0	0	**.96**	.04	0	0	.13	**.87**
	(10,50)	**.93**	.07	0	0	0	**1**	0	0	0	0	**.93**	.07	0	0	0	**1**
	(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

Table 4A: MSEs×1000: static panels, $\rho = 0$

True Model	Adopted Model	M1	M2	M3	M4	Selected by CV
M1	N=10 T=5	**8.03**	14.56	11.41	29.77	**9.83**
	N=50 T=5	**1.45**	2.82	2.05	5.23	**1.61**
	N=10 T=10	**3.79**	6.32	5.83	13.16	**4.16**
	N=50 T=10	**0.71**	1.23	1.02	2.31	**0.73**
	N=10 T=50	**0.68**	1.03	0.99	2.16	**0.69**
	N=50 T=50	**0.14**	0.21	0.21	0.41	**0.14**
M2	N=10 T=5	165.85	**14.56**	300.24	29.77	**36.35**
	N=50 T=5	148.68	**2.82**	262.69	5.23	**5.11**
	N=10 T=10	145.62	**6.32**	295.26	13.16	**6.99**
	N=50 T=10	130.7	**1.23**	259.24	2.31	**1.31**
	N=10 T=50	128.66	**1.03**	287.75	2.16	**1.03**
	N=50 T=50	117.76	**0.21**	258.94	0.41	**0.21**
M3	N=10 T=5	94.25	250.15	**11.41**	29.77	**14.62**
	N=50 T=5	82.01	223.07	**2.05**	5.23	**2.05**
	N=10 T=10	102.06	245.89	**5.83**	13.16	**6.66**
	N=50 T=10	94.97	235.31	**1.02**	2.31	**1.02**
	N=10 T=50	109.93	249.80	**0.99**	2.16	**1.01**
	N=50 T=50	107.42	246.93	**0.21**	0.41	**0.21**
M4	N=10 T=5	427.81	250.15	300.24	**29.77**	**103.77**
	N=50 T=5	404.04	223.07	262.69	**5.23**	**9.38**
	N=10 T=10	440.18	245.89	295.26	**13.16**	**17.93**
	N=50 T=10	422.87	235.31	259.24	**2.31**	**2.31**
	N=10 T=50	448.25	249.80	287.75	**2.16**	**2.16**
	N=50 T=50	441.74	246.93	258.94	**0.41**	**0.41**

Table 4B: Coverages and length of 95% CIs: static panels, $\rho = 0$

		Coverages					Length				
True Model	Adopted Model	M1	M2	M3	M4	CV	M1	M2	M3	M4	CV
M1	N=10 T=5	**0.93**	0.90	0.92	0.88	**0.91**	**0.32**	0.39	0.37	0.54	**0.32**
	N=50 T=5	**0.96**	0.91	0.94	0.91	**0.94**	**0.15**	0.18	0.17	0.25	**0.15**
	N=10 T=10	**0.93**	0.92	0.92	0.90	**0.92**	**0.23**	0.28	0.26	0.38	**0.23**
	N=50 T=10	**0.96**	0.93	0.95	0.94	**0.96**	**0.10**	0.13	0.12	0.17	**0.10**
	N=10 T=50	**0.94**	0.95	0.94	0.94	**0.94**	**0.10**	0.12	0.12	0.17	**0.10**
	N=50 T=50	**0.95**	0.95	0.95	0.95	**0.95**	**0.05**	0.06	0.05	0.08	**0.05**
M2	N=10 T=5	0.13	**0.90**	0.07	0.88	**0.81**	0.40	**0.39**	0.45	0.54	**0.40**
	N=50 T=5	0.00	**0.91**	0.00	0.91	**0.89**	0.19	**0.18**	0.21	0.25	**0.19**
	N=10 T=10	0.05	**0.92**	0.02	0.90	**0.91**	0.29	**0.28**	0.32	0.38	**0.28**
	N=50 T=10	0.00	**0.93**	0.00	0.94	**0.93**	0.13	**0.13**	0.15	0.17	**0.13**
	N=10 T=50	0.00	**0.95**	0.00	0.94	**0.95**	0.13	**0.12**	0.14	0.17	**0.12**
	N=50 T=50	0.00	**0.95**	0.00	0.95	**0.95**	0.06	**0.06**	0.07	0.08	**0.06**
M3	N=10 T=5	0.34	0.17	**0.92**	0.88	**0.89**	0.38	0.46	**0.37**	0.54	**0.37**
	N=50 T=5	0.11	0.04	**0.94**	0.91	**0.94**	0.18	0.21	**0.17**	0.25	**0.17**
	N=10 T=10	0.09	0.03	**0.92**	0.90	**0.91**	0.28	0.33	**0.26**	0.38	**0.27**
	N=50 T=10	0.01	0.00	**0.95**	0.94	**0.95**	0.13	0.15	**0.12**	0.17	**0.12**
	N=10 T=50	0.00	0.00	**0.94**	0.94	**0.94**	0.13	0.15	**0.12**	0.17	**0.12**
	N=50 T=50	0.00	0.00	**0.95**	0.95	**0.95**	0.06	0.07	**0.05**	0.08	**0.05**
M4	N=10 T=5	0.01	0.17	0.07	**0.88**	**0.66**	0.41	0.46	0.45	**0.54**	**0.49**
	N=50 T=5	0.00	0.04	0.00	**0.91**	**0.89**	0.19	0.21	0.21	**0.25**	**0.24**
	N=10 T=10	0.00	0.03	0.02	**0.90**	**0.88**	0.29	0.33	0.32	**0.38**	**0.38**
	N=50 T=10	0.00	0.00	0.00	**0.94**	**0.94**	0.13	0.15	0.15	**0.17**	**0.17**
	N=10 T=50	0.00	0.00	0.00	**0.94**	**0.94**	0.13	0.15	0.14	**0.17**	**0.17**
	N=50 T=50	0.00	0.00	0.00	**0.95**	**0.95**	0.06	0.07	0.07	**0.08**	**0.08**

Table 4C: MSEs×1000: dynamic panels without exogenous regressors, $\beta = 3/4$

		Non-bias correction					Bias correction				
	Adopted M	M1	M2	M3	M4	CV	M1	M2	M3	M4	CV
True M											
M1	N=10 T=5	**11.78**	208.92	11.13	214.24	**124.20**	**12.47**	88.12	11.64	98.40	**55.41**
	N=50 T=5	**1.90**	174.56	1.87	174.77	**51.53**	**1.96**	19.85	1.93	19.96	**8.95**
	N=10 T=10	**5.55**	57.34	5.40	58.74	**30.10**	**5.41**	25.20	5.52	28.49	**13.30**
	N=50 T=10	**0.95**	45.78	0.94	45.91	**6.53**	**0.94**	4.69	0.95	4.80	**1.71**
	N=10 T=50	**0.84**	2.48	0.92	2.64	**1.25**	**0.85**	1.51	0.94	1.69	**0.95**
	N=50 T=50	**0.17**	1.61	0.17	1.62	**0.17**	**0.17**	0.32	0.17	0.32	**0.17**
M2	N=10 T=5	47.88	**208.92**	49.05	214.24	**200.23**	47.28	**88.12**	48.72	98.40	**84.22**
	N=50 T=5	47.65	**174.56**	47.88	174.77	**167.70**	47.51	**19.85**	47.80	19.96	**23.21**
	N=10 T=10	46.82	**57.34**	48.10	58.74	**60.86**	46.54	**25.20**	47.92	28.49	**24.83**
	N=50 T=10	47.76	**45.78**	48.00	45.91	**46.19**	47.71	**4.69**	47.98	4.80	**4.97**
	N=10 T=50	46.77	**2.48**	48.04	2.64	**2.48**	46.86	**1.51**	48.11	1.69	**1.51**
	N=50 T=50	47.74	**1.61**	47.98	1.62	**1.61**	47.75	**0.32**	47.99	0.32	**0.32**
M3	N=10 T=5	38.70	261.27	**11.13**	214.24	**116.90**	62.42	201.05	**11.64**	98.40	**55.80**
	N=50 T=5	23.14	235.57	**1.87**	174.77	**49.00**	40.55	142.49	**1.93**	19.96	**8.60**
	N=10 T=10	20.31	79.86	**5.40**	58.74	**28.02**	24.37	56.13	**5.52**	28.49	**13.06**
	N=50 T=10	13.41	72.01	**0.94**	45.91	**6.38**	17.73	39.64	**0.95**	4.80	**1.72**
	N=10 T=50	3.16	5.09	**0.92**	2.64	**1.34**	3.85	4.59	**0.94**	1.69	**1.05**
	N=50 T=50	2.41	4.35	**0.17**	1.62	**0.17**	3.02	3.56	**0.17**	0.32	**0.17**
M4	N=10 T=5	37.69	261.27	49.05	**214.24**	**192.47**	35.79	201.05	48.72	**98.40**	**89.18**
	N=50 T=5	38.03	235.57	47.88	**174.77**	**165.80**	36.60	142.49	47.80	**19.96**	**23.82**
	N=10 T=10	36.30	79.86	48.10	**58.74**	**63.45**	36.09	56.13	47.92	**28.49**	**28.45**
	N=50 T=10	37.51	72.01	48.00	**45.91**	**46.21**	37.25	39.64	47.98	**4.80**	**5.10**
	N=10 T=50	36.75	5.09	48.04	**2.64**	**2.64**	37.13	4.59	48.11	**1.69**	**1.69**
	N=50 T=50	37.62	4.35	47.98	**1.62**	**1.62**	37.87	3.56	47.99	**0.32**	**0.32**

Table 4D: Coverages and length of 95% CIs (bias-corrected): dynamic panels without exogenous regressors, $\beta = 3/4$

		Coverages					Length				
	Adopted Model	M1	M2	M3	M4	CV	M1	M2	M 3	M4	CV
True Model											
M1	N=10 T=5	**0.91**	0.70	0.91	0.66	**0.75**	**0.37**	0.62	0.36	0.62	**0.45**
	N=50 T=5	**0.94**	0.67	0.93	0.66	**0.83**	**0.16**	0.28	0.16	0.28	**0.19**
	N=10 T=10	**0.93**	0.72	0.91	0.70	**0.84**	**0.26**	0.36	0.26	0.36	**0.29**
	N=50 T=10	**0.94**	0.77	0.94	0.76	**0.91**	**0.12**	0.16	0.12	0.16	**0.12**
	N=10 T=50	**0.95**	0.88	0.94	0.88	**0.94**	**0.12**	0.12	0.12	0.12	**0.12**
	N=50 T=50	**0.96**	0.88	0.96	0.89	**0.96**	**0.05**	0.06	0.05	0.06	**0.05**
M2	N=10 T=5	0.01	**0.70**	0.01	0.66	**0.50**	0.14	**0.62**	0.13	0.62	**0.49**
	N=50 T=5	0.00	**0.67**	0.00	0.66	**0.59**	0.06	**0.28**	0.06	0.28	**0.26**
	N=10 T=10	0.00	**0.72**	0.00	0.70	**0.68**	0.10	**0.36**	0.09	0.36	**0.34**
	N=50 T=10	0.00	**0.77**	0.00	0.76	**0.76**	0.04	**0.16**	0.04	0.16	**0.16**
	N=10 T=50	0.00	**0.88**	0.00	0.88	**0.88**	0.04	**0.12**	0.04	0.12	**0.12**
	N=50 T=50	0.00	**0.88**	0.00	0.89	**0.88**	0.02	**0.06**	0.02	0.06	**0.06**
M3	N=10 T=5	0.57	0.50	**0.91**	0.66	**0.77**	0.42	0.62	**0.36**	0.63	**0.44**
	N=50 T=5	0.36	0.29	**0.93**	0.67	**0.83**	0.19	0.28	**0.16**	0.28	**0.19**
	N=10 T=10	0.64	0.49	**0.91**	0.70	**0.84**	0.29	0.36	**0.26**	0.36	**0.28**
	N=50 T=10	0.30	0.27	**0.94**	0.76	**0.91**	0.13	0.16	**0.12**	0.16	**0.12**
	N=10 T=50	0.68	0.66	**0.94**	0.88	**0.93**	0.12	0.12	**0.12**	0.12	**0.12**
	N=50 T=50	0.36	0.36	**0.96**	0.89	**0.96**	0.05	0.06	**0.05**	0.06	**0.05**
M4	N=10 T=5	0.12	0.50	0.01	**0.66**	**0.42**	0.18	0.62	0.13	**0.63**	**0.45**
	N=50 T=5	0.01	0.29	0.00	**0.67**	**0.57**	0.08	0.28	0.06	**0.28**	**0.25**
	N=10 T=10	0.03	0.49	0.00	**0.70**	**0.63**	0.13	0.36	0.09	**0.36**	**0.32**
	N=50 T=10	0.00	0.27	0.00	**0.76**	**0.76**	0.06	0.16	0.04	**0.16**	**0.16**
	N=10 T=50	0.00	0.66	0.00	**0.88**	**0.88**	0.06	0.12	0.04	**0.12**	**0.12**
	N=50 T=50	0.00	0.36	0.00	**0.89**	**0.89**	0.03	0.06	0.02	**0.06**	**0.06**

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

	Model selection						Inference for the coefficient on the "probability of arrest"		
Model	AIC	BIC	BIC_2	CV	CV^*	CV^{**}	Estimates	95% CI^1	95% CI^2
Model 1	-2.121	-2.001	-2.125	0.124	0.094	0.028	-0.530	[-0.655, -0.406]	[-0.785, -0.276]
Model 2	-3.773	-3.025	-3.796	0.025	0.023	0.026	-0.385	[-0.473, -0.297]	[-0.500, -0.270]
Model 3	-2.124	-1.962	-2.129	0.124	0.094	0.027	-0.521	[-0.646, -0.396]	[-0.778, -0.264]
Model 4	-3.823	-3.032	-3.847	0.024	0.022	0.025	-0.355	[-0.441, -0.269]	[-0.470, -0.240]
Selected	M4	M4	M4	M4	M4	M4	-0.355	[-0.441, -0.269]	[-0.470, -0.240]

Notes: CI^1 and CI^2 stand for the CIs based on the non-clustered and clustered standard errors, respectively.

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

					Inference for the coefficent on the "GDP growth"			
	Model selection				Non-bias correction		Bias correction	
Model	AIC	BIC	BIC_2	CV	Estimates	95% CI	Estimates	95% CI
Model 1	2.547	2.576	2.547	12.844	0.190	[0.108, 0.273]	0.192	[0.107, 0.277]
Model 2	2.505	2.843	2.498	12.459	0.188	[0.088, 0.288]	0.178	[0.074, 0.281]
Model 3	2.555	2.663	2.553	12.953	0.160	[0.072, 0.248]	0.163	[0.073, 0.253]
Model 4	2.512	2.929	2.504	12.584	0.149	[0.039, 0.258]	0.146	[0.031, 0.262]
Selected	M2	M1	M2	M2	0.188	[0.088, 0.288]	0.178	[0.074, 0.281]

Table 7: Application III: Guns and crime in the U.S. (N=51, T=23, k=9)

	Model selection						Inference for the coefficient of the "shall issue"		
Model	AIC	BIC	BIC_2	CV	CV^*	CV^{**}	Estimates	95% CI^1	95% CI^2
log (violent crime rate)									
M1	-1.6911	-1.6522	-1.6914	0.1860	0.0165	0.0073	-0.368	[-0.436, -0.301]	[-0.589, -0.148]
M2	-3.6072	-3.3523	-3.6094	0.0274	0.0080	0.0072	-0.046	[-0.084, -0.008]	[-0.127, 0.035]
M3	-1.7198	-1.5859	-1.7210	0.1816	0.0140	0.0061	-0.288	[-0.359, -0.217]	[-0.526, -0.050]
M4	-3.8653	-3.5154	-3.8684	0.0211	0.0063	0.0059	-0.028	[-0.065, 0.009]	[-0.106, 0.050]
Selected	M4	M4	M4	M4	M4	M4	-0.028	[-0.065, 0.009]	[-0.106, 0.050]
log (murder rate)									
M1	-1.6202	-1.5813	-1.6205	0.1991	0.1234	0.0560	-0.313	[-0.383, -0.244]	[-0.505, -0.122]
M2	-2.9845	-2.7296	-2.9867	0.0510	0.0457	0.0452	-0.061	[-0.113, -0.009]	[-0.132, 0.011]
M3	-1.7012	-1.5673	-1.7024	0.1844	0.1144	0.0550	0.198	[-0.267, -0.130]	[-0.385, -0.012]
M4	-3.1243	-2.7744	-3.1274	0.0443	0.0413	0.0421	-0.015	[-0.066, 0.037]	[-0.088, 0.058]
Selected	M4	M4	M4	M4	M4	M4	-0.015	[-0.066, 0.037]	[-0.088, 0.058]
log (robbery rate)									
M1	-0.9853	-0.9464	-0.9856	0.3748	0.0375	0.0164	-0.529	[-0.628, -0.429]	[-0.840, -0.218]
M2	-3.0239	-2.7690	-3.0261	0.0490	0.0167	0.0156	-0.008	[-0.058, 0.043]	[-0.115, 0.099]
M3	-1.1079	-0.9740	-1.1091	0.3338	0.0305	0.0137	-0.341	[-0.436, -0.246]	[-0.642, -0.040]
M4	-3.2181	-2.8682	-3.2212	0.0403	0.0135	0.0130	0.027	[-0.021, 0.075]	[-0.073, 0.127]
Selected	M4	M4	M4	M4	M4	M4	0.027	[-0.021, 0.075]	[-0.073, 0.127]

Notes: CI^1 and CI^2 stand for the CIs based on the non-clustered and clustered standard errors, respectively.

Supplementary Appendix to
"Determining Individual or Time Effects in Panel Data Models"
(NOT for Publication)

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

This supplement is composed of three parts. Section B contains the proofs of the technical lemmas in the above paper. Section C provides some primitive conditions to verify Assumptions A.2(iii)-(iv) and A.4-A.5 in the paper. Section D discusses how to choose p in our modified jackknife and contains additional simulation results.

B Proofs of the Technical Lemmas

Proof of Lemma A.1. 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). ■

Proof of Lemma A.2. By Lemma A.1,

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

Noting that $D_1'D_2=0$, we have $\left(D'D\right)^{-1} = \begin{pmatrix} \left(D_1'D_1\right)^{-1} & \\ & \left(D_2'D_2\right)^{-1} \end{pmatrix}$, and $X'D\left(D'D\right)^{-1} = X'(D_1\left(D_1'D_1\right)^{-1},$ $D_2\left(D_2'D_2\right)^{-1}) = (B_1, B_2)$. Combining the above results yields the desired result. ■

Proof of Lemma A.3. (i) Noting that $D_\alpha'D_\alpha = T\left(I_{N-1} + \iota_{N-1}\iota_{N-1}'\right)$, we have

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

and

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

By straightforward but tedious algebra we can show that

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

The result then follows from Assumption A.1(iii).

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

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

(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(iii). ■

Proof of Lemma A.4. (i) Following the proof of Lemma A.3(i), we have

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

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

(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). ■

Proof of Lemma A.5. (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(iii)-(iv).

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

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

where $B_\alpha = X' D_\alpha (D_\alpha' D_\alpha)^{-1}$. As in (i), we can show that $J_{2NT,1} = O_P((NT)^{-1})$ by Assumption A.1(iii)-(iv). By Lemma A.4(i) and Assumptions A.1(iii)-(iv) and A.2(iii) and using $X_{D_\alpha}^* = (X' M_{D_\alpha} X)^{-1}$,

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

By Lemma A.3(i),

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

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

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

(iv) By Lemma A.2

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

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(iii)-(iv),

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

By Lemmas A.3(i) and A.4(i) and Assumption A.1(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 Assumption A.1(iii),

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

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

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

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

Proof of Lemma A.6. (i) $\max_{i,t}h_{it}^{(1)}=\max_{i,t}x_{it}'\left(X'X\right)^{-1}x_{it}\le\left[\lambda_{\min}\left((NT)^{-1}X'X\right)\right]^{-1}(NT)^{-1}\max_{i,t}\|x_{it}\|^2=O_P((NT)^{-1/2})$ by Assumption A.1(ii) and (iv).

(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

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

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 (B.1), imply that

$$d'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}d_{\alpha,it}=T^{-1}d'_{\alpha,it}\left(I_{N-1}-\frac{1}{N}\iota_{N-1}\iota'_{N-1}\right)d_{\alpha,it}=T^{-1}\frac{N-1}{N}\text{ for all }i,t. \tag{B.3}$$

Next, notice that

$$\begin{aligned}\max_{i,t}\left(x_{it}-B_\alpha d_{\alpha,it}\right)'X^*_{D_\alpha}\left(x_{it}-B_\alpha d_{\alpha,it}\right) &\leq \epsilon_{1NT}\max_{i,t}\frac{1}{NT}\left(x_{it}-B_\alpha d_{\alpha,it}\right)'\left(x_{it}-B_\alpha d_{\alpha,it}\right)\\ &\leq \epsilon_{1NT}\max_{i,t}\frac{2}{NT}\left\{\|x_{it}\|^2+d'_{\alpha,it}B'_\alpha B_\alpha d_{\alpha,it}\right\},\end{aligned}$$

where $\epsilon_{1NT}=\left[\lambda_{\min}\left(\frac{1}{NT}X'M_{D_\alpha}X\right)\right]^{-1}=O_P(1)$ by Assumption A.1(iv). By Assumption A.1(ii) and Markov inequality,

$$\max_{i,t}\frac{1}{NT}\|x_{it}\|^2=\frac{1}{NT}O_P((NT)^{1/2})=O_P((NT)^{-1/2}).$$

For $\frac{1}{NT}d'_{\alpha,it}B'_\alpha B_\alpha d_{\alpha,it}$, we have

$$\begin{aligned}&\max_{i,t}\frac{1}{NT}d'_{\alpha,it}B'_\alpha B_\alpha d_{\alpha,it}\\ =\;&\max_{i,t}\frac{1}{NT}\mathrm{tr}\left\{d'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1/2}\left[\left(D'_\alpha D_\alpha\right)^{-1/2}D'_\alpha XX'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1/2}\right]\left(D'_\alpha D_\alpha\right)^{-1/2}d_{\alpha,it}\right\}\\ \leq\;&\max_{i,t}d'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}d_{\alpha,it}\frac{1}{NT}\mathrm{tr}\left(\left(D'_\alpha D_\alpha\right)^{-1/2}D'_\alpha XX'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1/2}\right)\\ =\;&T^{-1}\frac{N-1}{N}\frac{1}{NT}\mathrm{tr}\left(XX'D_\alpha\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha\right)\\ \leq\;&T^{-1}\frac{N-1}{N}\frac{1}{NT}\mathrm{tr}\left(X'X\right)=O_P\left(T^{-1}\right),\end{aligned}$$

where the last inequality follows from the fact that $D_\alpha\left(D'_\alpha D_\alpha\right)^{-1}D'_\alpha$ is a projection matrix with maximum eigenvalue 1. It follows that $h^{(2)}_{it}=T^{-1}\frac{N-1}{N}+\left(x_{it}-B_\alpha d_{\alpha,it}\right)'X^*_{D_\alpha}\left(x_{it}-B_\alpha d_{\alpha,it}\right)$ and $\max_{i,t}h^{(2)}_{it}=O_P((NT)^{-1/2}+T^{-1})$.

(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

$$h^{(3)}_{it}=d'_{\lambda,it}\left(D'_\lambda D_\lambda\right)^{-1}d_{\lambda,it}+\left(x_{it}-B_\lambda d_{\lambda,it}\right)'X^*_{D_\lambda}\left(x_{it}-B_\lambda d_{\lambda,it}\right) \tag{B.4}$$

and

$$d'_{\alpha,it}\left(I_{T-1}-\frac{1}{T}\iota_{T-1}\iota'_{T-1}\right)d_{\alpha,it}=\frac{T-1}{T}\text{ for all }i,t. \tag{B.5}$$

Noting that

$$D'_\lambda D_\lambda=N\left(I_{T-1}+\iota_{T-1}\iota'_{T-1}\right)\text{ and }\left(D'_\lambda D_\lambda\right)^{-1}=N^{-1}\left(I_{T-1}-\frac{1}{T}\iota_{T-1}\iota'_{T-1}\right), \tag{B.6}$$

we have

$$d'_{\lambda,it}\left(D'_\lambda D_\lambda\right)^{-1}d_{\lambda,it}=N^{-1}\frac{T-1}{T}. \tag{B.7}$$

In addition, following the arguments as used in the analysis of $\left(x_{it}-B_\alpha d_{\alpha,it}\right)' X^*_{D_\alpha}\left(x_{it}-B_\alpha d_{\alpha,it}\right)$ and (B.6), we have

$$\max_{i,t}\left(x_{it}-B_\lambda d_{\lambda,it}\right)' X^*_{D_\alpha}\left(x_{it}-B_\lambda d_{\lambda,it}\right)\leq \epsilon_{2NT}\max_{i,t}\frac{2}{NT}\left(\|x_{it}\|^2+d'_{\lambda,it}B'_\lambda B_\lambda d_{\lambda,it}\right)=O_P((NT)^{-1/2}+N^{-1}),$$

where $\epsilon_{2NT}=\left[\lambda_{\min}\left(\frac{1}{NT}X'M_{D_\lambda}X\right)\right]^{-1}=O_P\left(1\right)$ by Assumption A.1(iv). It follows that $h^{(3)}_{it}=N^{-1}\frac{T-1}{T}+\left(x_{it}-B_\lambda d_{\lambda,it}\right)' X^*_{D_\lambda}\left(x_{it}-B_\lambda d_{\lambda,it}\right)$ and $\max_{i,t}h^{(3)}_{it}=O_P((NT)^{-1/2}+N^{-1})$.

(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

$$h^{(4)}_{it}=d'_{\alpha\lambda,it}\left(D'_{\alpha\lambda}D_{\alpha\lambda}\right)^{-1}d_{\alpha\lambda,it}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)' X^*_{D_{\alpha\lambda}}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right). \tag{B.8}$$

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

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

Then

$$\begin{aligned} d'_{\alpha\lambda,it}\left(D'_{\alpha\lambda}D_{\alpha\lambda}\right)^{-1}d_{\alpha\lambda,it} &= d'_{\alpha,it}\left(D'_\alpha D_\alpha\right)^{-1}d_{\alpha,Nt}+d'_{\lambda,it}\left(D'_\lambda D_\lambda\right)^{-1}d_{\lambda,it} \\ &= T^{-1}\frac{N-1}{N}+N^{-1}\frac{T-1}{T}\text{ for all } i,t. \end{aligned} \tag{B.9}$$

In addition, following the arguments as used in the analysis of $\left(x_{it}-B_\alpha d_{\alpha,it}\right)' X^*_{D_\alpha}\left(x_{it}-B_\alpha d_{\alpha,it}\right)$ and (B.9), we can show that

$$\begin{aligned} \max_{i,t}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)' X^*_{D_{\alpha\lambda}}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right) &\leq \epsilon_{3NT}\max_{i,t}\frac{2}{NT}\left\{\|x_{it}\|^2+d'_{\alpha\lambda,it}B'_{\alpha\lambda}B_{\alpha\lambda}d_{\alpha\lambda,it}\right\} \\ &= O_P((NT)^{-1/2}+N^{-1}+T^{-1})=O_P(N^{-1}+T^{-1}), \end{aligned}$$

where $\epsilon_{3NT}=\left[\lambda_{\min}\left(\frac{1}{NT}X'M_{D_{\alpha\lambda}}X\right)\right]^{-1}=O_P\left(1\right)$ by Assumption A.1(iv). It follows that $h^{(4)}_{it}=T^{-1}\frac{N-1}{N}+N^{-1}\frac{T-1}{T}+\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)' X^*_{D_{\alpha\lambda}}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)$ and $\max_{i,t}h^{(4)}_{it}=O_P(N^{-1}+T^{-1})$.

(v) Note that

$$\begin{aligned} &\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(x_{it}-B_\alpha d_{\alpha,it}\right)' X^*_{D_\alpha}\left(x_{it}-B_\alpha d_{\alpha,it}\right)u^2_{it} \\ \leq\ &\frac{2\epsilon_{1NT}}{NT}\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T x'_{it}x_{it}u^2_{it}+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T d'_{\alpha,it}B'_\alpha B_\alpha d_{\alpha,it}u^2_{it}\right\}\equiv\frac{2\epsilon_{1NT}}{NT}\left(II_1+II_2\right),\text{ say.} \end{aligned}$$

Note that $II_1=O_P(1)$ by Assumption A.1(i)-(ii) and Markov and Jensen inequalities. For II_2, observing that $X'D_\alpha\left(D_\alpha'D_\alpha\right)^{-1}d_{\alpha,it}=\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}$, we have by Assumption A.1(i)-(ii)

$$\begin{aligned}
II_2 &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{\alpha,it}'\left(D_\alpha'D_\alpha\right)^{-1}D_\alpha'XX'D_\alpha\left(D_\alpha'D_\alpha\right)^{-1}d_{\alpha,it}u_{it}^2\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)'\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)u_{it}^2\\
&\leq \left\{\frac{1}{N}\sum_{i=1}^{N}\left\|\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right\|^4\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}u_{it}^4\right\}^{1/2}=O_P\left(1\right)
\end{aligned}$$

because we can readily show that $\frac{1}{N}\sum_{i=1}^{N}\left\|\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right\|^4\leq\frac{8}{N}\sum_{i=1}^{N}\left\|\bar{x}_{i\cdot}\right\|^4+8\left\|\bar{x}_{\cdot\cdot}\right\|^4=O_P\left(1\right)$ under Assumption A.1(ii). It follows that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-B_\alpha d_{\alpha,it}\right)'X_{D_\alpha}^*\left(x_{it}-B_\alpha d_{\alpha,it}\right)u_{it}^2=O_P((NT)^{-1})$

(vi) The proof of (vi) is analogous to that of (v). The major difference is now we need to apply the fact that $X'D_\lambda\left(D_\lambda'D_\lambda\right)^{-1}d_{\lambda,it}=\bar{x}_{\cdot t}-\bar{x}_{\cdot\cdot}$.

(vii) Note that

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)'X_{D_{\alpha\lambda}}^*\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)u_{it}^2\\
\leq\ &\frac{2\epsilon_{3NT}}{NT}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}x_{it}'x_{it}u_{it}^2+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{\alpha\lambda,it}'B_{\alpha\lambda}'B_{\alpha\lambda}d_{\alpha\lambda,it}u_{it}^2\right\}\equiv\frac{2\epsilon_{3NT}}{NT}\left(II_1+II_3\right),\text{ say.}
\end{aligned}$$

We know that $II_1=O_P(1)$. For II_3, observing that $X'D_{\alpha\lambda}\left(D_{\alpha\lambda}'D_{\alpha\lambda}\right)^{-1}d_{\alpha\lambda,it}=\bar{x}_{i\cdot}+\bar{x}_{\cdot t}-2\bar{x}_{\cdot\cdot}$, we have by Assumption A.1(i)-(ii)

$$\begin{aligned}
II_2 &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}d_{\alpha\lambda,it}'\left(D_{\alpha\lambda}'D_{\alpha\lambda}\right)^{-1}D_{\alpha\lambda}'XX'D_{\alpha\lambda}\left(D_{\alpha\lambda}'D_{\alpha\lambda}\right)^{-1}d_{\alpha\lambda,it}u_{it}^2\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\bar{x}_{i\cdot}+\bar{x}_{\cdot t}-2\bar{x}_{\cdot\cdot}\right)'\left(\bar{x}_{i\cdot}+\bar{x}_{\cdot t}-2\bar{x}_{\cdot\cdot}\right)u_{it}^2\\
&\leq \frac{3}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\left\|\bar{x}_{i\cdot}\right\|^2+\left\|\bar{x}_{\cdot t}\right\|^2+4\left\|\bar{x}_{\cdot\cdot}\right\|^2\right)u_{it}^2=O_P\left(1\right).
\end{aligned}$$

Consequently, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)'X_{D_{\alpha\lambda}}^*\left(x_{it}-B_{\alpha\lambda}d_{\alpha\lambda,it}\right)u_{it}^2=O_P((NT)^{-1})$.■

Proof of Lemma A.7. Let $\|A\|_{\rm sp}$ denotes the spectral norm of A. Note that

$$\begin{aligned}
\hat{\boldsymbol{\rho}}-\boldsymbol{\rho} &= \left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\boldsymbol{\rho}\\
&= \left[\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\right]\hat{\mathbf{Z}}'\hat{\mathbf{U}}+\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\left(\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right)+\left[\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U}-\boldsymbol{\rho}\right]. \qquad \text{(B.10)}
\end{aligned}$$

It suffices to show that (i) $\frac{1}{NT_p}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{Z}}-\mathbf{Z}'\mathbf{Z}\right\|_{\rm sp}=O_P(p\delta_{NT})$, (ii) $\left\|\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}-\Gamma_p\right\|_{\rm sp}=o_P(1)$, (iii) $\frac{1}{NT_p}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right\|$ $=O_P(p^{1/2}\delta_{NT})$, and (iv) $\left\|\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U}-\boldsymbol{\rho}\right\|=O_P(p^{1/2}\delta_{NT})$. To see this, note that the last term in (B.10) is bounded the desired probability order by (iv). The second term in (B.10) is bounded above by

$$
\begin{aligned}
\left\|(\mathbf{Z}'\mathbf{Z})^{-1}\left(\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right)\right\| &\leq \left\|(\mathbf{Z}'\mathbf{Z})^{-1}\right\|_{\mathrm{sp}}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right\| \\
&= \left[\lambda_{\min}\left(\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}\right)\right]^{-1}\frac{1}{NT_p}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right\| \\
&\leq \left[\lambda_{\min}(\Gamma_p)+\left\|\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}-\Gamma_p\right\|\right]^{-1}\frac{1}{NT_p}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right\| \\
&= O_P(1)\,O_P(p^{1/2}\delta_{NT})=O_P(p^{1/2}\delta_{NT}),
\end{aligned}
$$

where the second inequality follows from the eigenvalue stability inequality, and the last line follows from the fact that $\lambda_{\min}(\Gamma_p)$ is bounded away from zero in probability. For the first term in (B.10), we have

$$
\begin{aligned}
\left\|\left[\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-(\mathbf{Z}'\mathbf{Z})^{-1}\right]\hat{\mathbf{Z}}'\hat{\mathbf{U}}\right\| &\leq \left\|\left(\frac{1}{NT_p}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-\left(\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}\right)^{-1}\right\|_{\mathrm{sp}}\frac{1}{NT_p}\left\{\|\mathbf{Z}'\mathbf{U}\|+\left\|\hat{\mathbf{Z}}'\hat{\mathbf{U}}-\mathbf{Z}'\mathbf{U}\right\|\right\} \\
&= O_P(p\delta_{NT})O_P(1+p^{1/2}\delta_{NT})=O_P(p\delta_{NT}),
\end{aligned}
$$

where we use the fact that by (i)-(iii),

$$
\begin{aligned}
\left\|\left(\frac{1}{NT_p}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}-\left(\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}\right)^{-1}\right\|_{\mathrm{sp}} &= \left\|\left(\frac{1}{NT_p}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)^{-1}\frac{1}{NT_p}\left(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}-\mathbf{Z}'\mathbf{Z}\right)\left(\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}\right)^{-1}\right\|_{\mathrm{sp}} \\
&\leq \left[\lambda_{\min}\left(\frac{1}{NT_p}\hat{\mathbf{Z}}'\hat{\mathbf{Z}}\right)\lambda_{\min}\left(\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}\right)\right]^{-1}\frac{1}{NT_p}\left\|\hat{\mathbf{Z}}'\hat{\mathbf{Z}}-\mathbf{Z}'\mathbf{Z}\right\|_{\mathrm{sp}} \\
&= O_P(1)\,O_P(p\delta_{NT}),
\end{aligned}
$$

and that $\frac{1}{NT_p}\|\mathbf{Z}'\mathbf{U}\|=O_P(1)$ under Assumption A.4. Next, we show (i)-(iv) in turn.

To show (i), 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

$$
\widehat{\ddot{u}}_{it}=\ddot{y}_{it}-\hat{\beta}^{*\prime}\ddot{x}_{it}^*=\ddot{u}_{it}-(\hat{\beta}^*-\beta^*)'\ddot{x}_{it}^*, \tag{B.11}
$$

Under Assumptions A.1(iii)-(iv) and A.2(iii)-(iv), we can readily show that $\hat{\beta}^*-\beta^*=O_P(\delta_{NT})$. Recall that $\underline{\ddot{u}}_{i,t} = (\ddot{u}_{it}, ..., \ddot{u}_{i,t-p+1})'$. Let $\widehat{\underline{\ddot{u}}}_{i,t} = (\widehat{\ddot{u}}_{i,t}, ...\widehat{\ddot{u}}_{i,t-p+1})'$. Then

$$\begin{aligned}
\frac{1}{NT_p}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}} - \mathbf{Z}'\mathbf{Z}) &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\underline{\ddot{u}}}_{i,t}\widehat{\underline{\ddot{u}}}'_{i,t} - \underline{\ddot{u}}_{i,t}\underline{\ddot{u}}'_{i,t}\right) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\underline{\ddot{u}}}_{i,t} - \underline{\ddot{u}}_{i,t}\right)\left(\widehat{\underline{\ddot{u}}}_{i,t} - \underline{\ddot{u}}_{i,t}\right)' + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left(\widehat{\underline{\ddot{u}}}_{i,t} - \underline{\ddot{u}}_{i,t}\right)\underline{\ddot{u}}'_{i,t} \\
&\quad + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t}\left(\widehat{\underline{\ddot{u}}}_{i,t} - \underline{\ddot{u}}_{i,t}\right)' \\
&\equiv \vartheta_1 + \vartheta_2 + \vartheta_3, \text{ say.}
\end{aligned}$$

Noting that $\widehat{\underline{\ddot{u}}}_{i,t} - \underline{\ddot{u}}_{i,t} = -\underline{\ddot{x}}^*_{it}(\hat{\beta}^* - \beta^*)$ where $\underline{\ddot{x}}^*_{it} = (\ddot{x}^*_{it}, ..., \ddot{x}^*_{i,t-p+1})'$, we have

$$\|\vartheta_1\| \leq \left\|\hat{\beta}^* - \beta^*\right\|^2 \left\|\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{x}}^*_{it}\underline{\ddot{x}}^{*\prime}_{it}\right\| = O_P(p\delta^2_{NT}).$$

For ϑ_2, we have

$$\begin{aligned}
\|\vartheta_2\|^2 &= \frac{1}{(NT_p)^2}\left\|\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{x}}^*_{it}(\hat{\beta}^* - \beta^*)\underline{\ddot{u}}'_{i,t}\right\|^2 \\
&= \frac{1}{(NT_p)^2}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\sum_{i_1=1}^{N}\sum_{t_1=p+1}^{T}\operatorname{tr}\left\{(\hat{\beta}^* - \beta^*)\underline{\ddot{u}}'_{i,t}\underline{\ddot{u}}_{i_1,t_1}(\hat{\beta}^* - \beta^*)'\underline{\ddot{x}}^{*\prime}_{i_1t_1}\underline{\ddot{x}}^*_{it}\right\} \\
&= (\hat{\beta}^* - \beta^*)'\Theta_2(\hat{\beta}^* - \beta^*),
\end{aligned}$$

where $\Theta_2 = \frac{1}{(NT_p)^2}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\sum_{i_1=1}^{N}\sum_{t_1=p+1}^{T}\underline{\ddot{u}}'_{i,t}\underline{\ddot{u}}_{i_1,t_1}\underline{\ddot{x}}^{*\prime}_{it}\underline{\ddot{x}}^*_{i_1t_1}$. Observe that under Assumption A.1(i)-(ii)

$$E\|\Theta_2\|_{\text{sp}} \leq E\operatorname{tr}(\Theta_2) = \frac{1}{(NT_p)^2}\sum_{j=0}^{p-1}\sum_{l=0}^{p-1}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\sum_{i_1=1}^{N}\sum_{t_1=p+1}^{T}E\left(\ddot{u}_{i,t-j}\ddot{u}_{i_1,t_1-j}\ddot{x}^{*\prime}_{i,t-l}\underline{\ddot{x}}^*_{i_1,t_1-l}\right) = O\left(p^2\right).$$

It follows that $\|\Theta_2\|_{\text{sp}} = O_P\left(p^2\right)$ and $\|\vartheta_2\| \leq \left\|\hat{\beta}^* - \beta^*\right\| \|\Theta_2\|^{1/2}_{\text{sp}} = O_P\left(p\delta_{NT}\right)$. Similarly, $\|\vartheta_3\| = \|\vartheta_2\| = O_P(p\delta_{NT})$. [When x_{it} is strictly exogenous, we can show that $E\operatorname{tr}(\Theta_2) = O\left(p^2\delta_{NT}\right)$ and then $\|\vartheta_3\| = \|\vartheta_2\| \leq \left\|\hat{\beta}^* - \beta^*\right\| \|\Theta_2\|^{1/2}_{\text{sp}} = O_P\left(p\delta^2_{NT}\right) = o_P(p^{1/2}\delta_{NT})$.] Then (i) follows.

To show (ii), we note that $\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z} = \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t-1}\underline{\ddot{u}}'_{i,t-1}$, where $\underline{\ddot{u}}_{i,t-1} = (\ddot{u}_{i,t-1}, ..., \ddot{u}_{i,t-p})'$. Noting that $\ddot{u}_{it} = u_{it} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t} + \bar{u}_{\cdot\cdot}$ for each (i,t), the (j,l) th element of $\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z}$ is given by

$$\begin{aligned}
\frac{1}{NT_p}[\mathbf{Z}'\mathbf{Z}]_{j,l} &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\ddot{u}_{i,t-j}\ddot{u}_{i,t-l} \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}(u_{i,t-j} - \bar{u}_{i\cdot})(u_{i,t-l} - \bar{u}_{i\cdot}) - \frac{1}{T_p}\sum_{t=p+1}^{T}(\bar{u}_{\cdot,t-j} - \bar{u}_{\cdot\cdot})(\bar{u}_{\cdot,t-l} - \bar{u}_{\cdot\cdot}) \\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}u_{i,t-j}u_{i,t-l} + [B_2]_{j,l} \equiv [B_1]_{j,l} + [B_2]_{j,l},
\end{aligned}$$

where $[B_2]_{j,l} = -\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} u_{i,t-j}\bar{u}_{i\cdot} - \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\bar{u}_{i\cdot}u_{i,t-l} + \frac{1}{N}\sum_{i=1}^{N}\bar{u}_{i\cdot}^2 - \frac{1}{T_p}\sum_{t=p+1}^{T}\bar{u}_{\cdot,t-j}\bar{u}_{\cdot,t-l}$ $+\frac{1}{T_p}\sum_{t=p+1}^{T}\bar{u}_{\cdot,t-j}\bar{u}_{\cdot\cdot} + \frac{1}{T_p}\sum_{t=p+1}^{T}\bar{u}_{\cdot\cdot}\bar{u}_{\cdot,t-l} - (\bar{u}_{\cdot\cdot})^2$, $[B_\ell]_{j,l}$ denotes the (j,l)th element of B_ℓ for $\ell = 1, 2$, and B_ℓ's are implicitly defined. It is easy to show that $\|B_2\| = O_P\left(p\delta_{NT}^{-1}\right)$. Consequently, we have

$$\left\|\frac{1}{NT_p}\mathbf{Z}'\mathbf{Z} - \Gamma_p\right\| = \left\|\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t-1}\underline{\ddot{u}}_{i,t-1}' - \Gamma_p\right\| \le \|B_2\| = O_P\left(p\delta_{NT}^{-1}\right) = o_P(1).$$

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

Lastly, we show (iv) Let $\bar{v}_{\cdot t} = \bar{u}_{\cdot t} - \boldsymbol{\rho}'\underline{\bar{u}}_{\cdot,t-1}$ where $\underline{\bar{u}}_{\cdot,t-1} = (\bar{u}_{\cdot,t-1}, ..., \bar{u}_{\cdot,t-p})'$. 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

$$\begin{aligned}
\ddot{u}_{it} - \boldsymbol{\rho}'\underline{\ddot{u}}_{i,t-1} &= \left(u_{it} - \boldsymbol{\rho}'\underline{u}_{i,t-1}\right) - \Phi(1)\,\bar{u}_{i\cdot} - \left(\bar{u}_{\cdot t} - \boldsymbol{\rho}'\underline{\bar{u}}_{\cdot,t-1}\right) + \Phi(1)\,\bar{u}_{\cdot\cdot} \\
&= v_{it} - \Phi(1)\,\bar{u}_{i\cdot} - \bar{v}_{\cdot t} + \Phi(1)\,\bar{u}_{\cdot\cdot}.
\end{aligned}$$

Then

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

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 can readily apply Assumptions A.1(i), (iii) and A.2(iii)-(iv) and show that $\left\|\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t-1}\bar{u}_{i\cdot}\right\|$ $= O_P\left(p^{1/2}T^{-1}\right)$. For example, the first element of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t-1}\bar{u}_{i\cdot}$ is

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

Similarly, we can show that $\left\|\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{\ddot{u}}_{i,t-1}v_{it}\right\| = \left\|\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\underline{u}_{i,t-1}v_{it}\right\| + O_P\left(p^{1/2}\delta_{NT}\right)$ $= O_P\left(p^{1/2}\delta_{NT}\right)$ under Assumptions A.1(iii), A.4(iii), and A.5(i)-(ii). Then we have $\left\|\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{U} - \boldsymbol{\rho}\right\| =$ $O_P(p^{1/2}\delta_{NT})$.■

Proof of Lemma A.8. (i) Noting that $\breve{x}_{it}^{(1)} = x_{it} - \underline{x}_{i,t-1}\boldsymbol{\rho} = \boldsymbol{\Phi}(L)\,x_{it} \equiv \breve{x}_{it}$ where $\underline{x}_{i,t-1} = (x_{i,t-1}, ..., x_{i,t-p})$, we can readily apply Assumptions A.1(iv)-(v) and A.4(iii) to show that

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

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

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

As in (i), we can show that $K_{2NT,1}=O_P((NT)^{-1})$ by Assumption A.1(iv)-(v). Observing that $\breve{d}_{\alpha,it}'\left(D_\alpha' D_\alpha\right)^{-1}$ $\times D_\alpha' X=(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})'$ for each (i,t), we have

$$
\begin{aligned}
\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}\breve{d}_{\alpha,it}'B_\alpha' &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}\left(d_{\alpha,it}-\underline{d}_{\alpha,it-1}\boldsymbol{\rho}\right)'\left(D_\alpha' D_\alpha\right)^{-1}D_\alpha' X \\
&= \frac{\Phi(1)}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}\left(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot}\right)'=O_P(T^{-1}+(NT)^{-1/2}).
\end{aligned}
$$

by Assumptions A.4(iii) and A.5(iii). It follows that

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

Similarly, we can show that

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

and

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

It follows that $K_{2NT}=\frac{\Phi(1)}{NT_{p+1}}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}\bar{u}_{i\cdot}+O_P((NT)^{-1}+T^{-2})$.

(iii) The proof is analogous to that of (ii). The major difference is that we need to use the expression $\breve{x}_{it}^{(3)}=x_{it}^{(3)}-\underline{x}_{i,t-1}^{(3)}\boldsymbol{\rho}=\left((x_{it}-\underline{x}_{i,t-1}\boldsymbol{\rho})',(d_{\lambda,it}-\underline{d}_{\lambda,it-1}\boldsymbol{\rho})'\right)'\equiv\left(\breve{x}_{it}',\breve{d}_{\lambda,it}'\right)'$ with $\underline{d}_{\lambda,it-1}=(d_{\lambda,it-1},...,d_{\lambda,it-p})$ and the fact that $d_{\lambda,it}'(D_\lambda'D_\lambda)^{-1}D_\lambda'X=(\bar{x}_{\cdot t}-\bar{x}_{\cdot\cdot})'$ for each (i,t) to obtain

$$
\begin{aligned}
\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}\breve{d}_{\lambda,it}'B_\lambda' &= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}\left(d_{\lambda,it}-\underline{d}_{\lambda,it-1}\boldsymbol{\rho}\right)'(D_\alpha'D_\alpha)^{-1}D_\alpha'X\\
&= \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}v_{it}\left[(\bar{x}_{\cdot t}-\bar{x}_{\cdot\cdot})'-\boldsymbol{\rho}'\left(\underline{\bar{x}}_{\cdot t-1}-\bar{x}_{\cdot\cdot}\iota_p'\right)'\right],
\end{aligned}
$$

where $\underline{\bar{x}}_{\cdot t-1}=(\bar{x}_{\cdot t-1},...,\bar{x}_{\cdot t-p})$. The dominant term then becomes

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

where $\underline{\bar{u}}_{\cdot t-1}=(\bar{u}_{\cdot t-1},...,\bar{u}_{\cdot t-p})'$.

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

Proof of Lemma A.9. (i) Noting that $\breve{x}_{it}^{(1)}=x_{it}-\underline{x}_{i,t-1}\boldsymbol{\rho}\equiv\breve{x}_{it}$, we can readily apply Assumptions A.1(iv)-(v) and A.4(i) to show that

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

(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

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

where $\zeta_1 = X_{D_\alpha}^{*\prime} X'U - X_{D_\alpha}^{*\prime} B_\alpha D_\alpha' U$ and $\zeta_2 = -B_\alpha' X_{D_\alpha}^{*\prime} X'U + (D_\alpha' D_\alpha)^{-1} D_\alpha' U + B_\alpha' X_{D_\alpha}^{*\prime} B_\alpha D_\alpha' U$. 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 $L_{2NT,2}$,

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

Noting that $\breve{d}_{\alpha,it}' B_\alpha' = \breve{d}_{\alpha,it}' (D_\alpha' D_\alpha)^{-1} D_\alpha' X = \Phi(1) (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})'$, we have

$$B_\alpha \frac{1}{NT_p} \sum_{i=1}^{N} \sum_{t=p+1}^{T} \breve{d}_{\alpha,it} \breve{d}_{\alpha,it}' B_\alpha' = \Phi(1)^2 \frac{1}{N} \sum_{i=1}^{N} (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot}) (\bar{x}_{i\cdot} - \bar{x}_{\cdot\cdot})' = O_P(1).$$

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

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

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

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

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

(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. ■

C Verification of Some Assumptions

In this section, we verify Assumptions A.2(iii)-(iv) and A.4-A.5 based on some primitive conditions.

C.1 Verification of the rate conditions in Assumption A.2(iii)-(iv)

In this subsection, we verify the rate conditions in Assumption A.2(iii)-(iv). Recall that we use C to denote a generic positive constant whose value can change across lines. Let $x_{it}^{*}=x_{it}-E(x_{it})$. To verify the rate conditions in Assumption A.2(iii)-(iv), we add the following assumptions.

Assumption A.2* (i) $\max_{1\leq i\leq N}E\left\Vert \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^{*}u_{is}\right\Vert ^{2}\leq C$;

(ii) $\max_{1\leq t\leq T}E\left\Vert \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{it}^{*}u_{jt}\right\Vert ^{2}\leq C$;

(iii) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\vert E\left(u_{it}u_{js}\right)\right\vert \leq C.$

The conditions of the above type are frequently assumed in the panel data literature to control weak serial and cross-section dependence; see, e.g., Bai and Ng (2002). Below we first show that Assumption A.2*, in conjunction with Assumption A.1(i)-(ii), is sufficient for Assumptions A.2(iii)-(iv), and then give more primitive conditions to ensure Assumption A.2*(i). Similar primitive conditions can ensure Assumption A.2*(ii) by relying upon some mixing conditions in random field to handle weak cross-sectional dependence.

First, we verify Assumption A.2(iii): $\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot}=O_{P}(T^{-1}+(NT)^{-1/2})$. Let $\bar{x}_{i\cdot}^{*}=\bar{x}_{i\cdot}-E(\bar{x}_{i\cdot})$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot}=\frac{1}{N}\sum_{i=1}^{N}E\left(\bar{x}_{i\cdot}\right)\bar{u}_{i\cdot}+\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}^{*}\bar{u}_{i\cdot}\equiv A_{1}+A_{2},\text{ say.}$$

For A_{1}, we have $A_{1}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\bar{x}_{i\cdot}\right)u_{it}$. Note that $E\left(A_{1}\right)=0$ and

$$\begin{aligned}E\left\Vert A_{1}\right\Vert ^{2} &= \frac{1}{N^{2}T^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\bar{x}_{i\cdot}\right)^{\prime}E\left(\bar{x}_{j\cdot}\right)E\left(u_{it}u_{js}\right)\\ &\leq \frac{C}{N^{2}T^{2}}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\vert E\left(u_{it}u_{js}\right)\right\vert =O((NT)^{-1}),\end{aligned}$$

where the last equality follows from Assumption A.2*(iii). Then $A_{1}=O_{P}((NT)^{-1/2})$ by Chebyshev inequality. For A_{2}, we apply the Jensen inequality and Assumption A.2*(i) to obtain

$$E\left[\left\Vert A_{2}\right\Vert ^{2}\right]=E\left[\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^{*}u_{is}\right)^{2}\right]\leq\frac{1}{N}\sum_{i=1}^{N}E\left[\left(\frac{1}{T^{2}}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^{*}u_{is}\right)^{2}\right]\leq\frac{C}{T^{2}}.$$

It follows that $A_{2}=O_{P}\left(T^{-1}\right)$ by Chebyshev inequality. In sum, we have shown $\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{u}_{i\cdot}=O_{P}(T^{-1}+(NT)^{-1/2})$.

Next, we verify Assumption A.2(iv): $\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{t\cdot}\bar{u}_{t\cdot}=O_{P}(N^{-1}+(NT)^{-1/2})$. Let $\bar{x}_{\cdot t}^{*}=\bar{x}_{\cdot t}-E(\bar{x}_{\cdot t})$. Then

$$\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{t\cdot}\bar{u}_{t\cdot}=\frac{1}{T}\sum_{t=1}^{T}E\left(\bar{x}_{\cdot t}\right)\bar{u}_{\cdot t}+\frac{1}{T}\sum_{t=1}^{T}\bar{x}_{\cdot t}^{*}\bar{u}_{\cdot t}\equiv A_{3}+A_{4},\text{ say.}$$

For A_3, we have $A_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} E\left(\bar{x}_{\cdot t}\right) u_{it}$. Note that $E\left(A_3\right) = 0$ and

$$\begin{aligned} E\left(\|A_3\|^2\right) &\leq \frac{1}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left(\bar{x}_{\cdot t}\right)' E\left(\bar{x}_{\cdot s}\right) E\left(u_{it}u_{js}\right) \\ &\leq \frac{C}{N^2T^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \left|E\left(u_{it}u_{js}\right)\right| = O((NT)^{-1}) \text{ by Assumption A.2*(iii).} \end{aligned}$$

Hence $A_3 = O_P\left((NT)^{-1/2}\right)$ by Chebyshev inequality. For A_4, we have by Jensen inequality and Assumption A.2*(ii)

$$E\left[\|A_4\|^2\right] = E\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{it}^* u_{jt}\right) \leq \frac{1}{T}\sum_{t=1}^{T} E\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{it}^* u_{jt}\right)^2 \leq \frac{C}{N^2}.$$

It follows that $A_4 = O_P\left(N^{-1}\right)$ by Chebyshev inequality. In sum, we have that $\frac{1}{T}\sum_{t=1}^{T} \bar{x}_{t\cdot}\bar{u}_{t\cdot} = O_P(N^{-1} + (NT)^{-1/2})$.

Now, we provide a set of sufficient primitive conditions for Assumption A.2*(i).

Assumption A.2* (i.a) $\max_{i,t} E\left\|x_{it}\right\|^{4(1+\delta)} < C$ and $\max_{i,t} E\left|u_{it}\right|^{4(1+\delta)} < \infty$ for some $\delta > 0$;

(i.b) For each $i = 1, ..., N$, $\{(x_{it}, u_{it}),\ t \geq 1\}$ is a strong stationary strong mixing process with mixing coefficients $\alpha_i(\cdot)$ such that $\max_i \sum_{\tau=1}^{T} \tau^2 \alpha_i(\tau)^{\delta/(1+\delta)} < C$ for some $C < \infty$.

Assumption A.2*(i.a) strengthens the moment conditions in Assumption A.1(i)-(ii) slightly for the application of Davydov inequality. Assumption A.2*(i.b) requires that $\{(x_{it}, u_{it}),\ t \geq 1\}$ be strong mixing. This condition is a standard condition assumed for dynamic panels when the individual effect is assumed to be fixed. For example, for a dynamic panel autoregressive process of order one (PAR(1)), it is strong mixing with the mixing coefficient $\alpha_i(\tau)$ decaying to zero at a rate proportional to $|\rho|^{\tau}$ as long as the autoregressive coefficient ρ is strictly less than 1 in absolute value. In this case, Assumption A.2*(i.b) is automatically satisfied for all $\rho \in (-1, 1)$. If the individual effect is random, then we can replace the strong mixing condition by the corresponding conditional mixing condition: $\{(x_{it}, u_{it}),\ t \geq 1\}$ is conditionally strong mixing with mixing coefficients $\alpha_i(\cdot)$ given the individual effect. See Prakasa Rao (2009) for the definition of conditional strong mixing, and Hahn and Kuersteiner (2011) and Su and Chen (2013) for the applications of conditional strong mixing in dynamic panels.

We now show that Assumption A.2*(i.a)-(i.b) is sufficient for Assumption A.2*(i). By Cauchy-Schwarz inequality,

$$\max_i E\left\|\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it}^* u_{is}\right\|^2 \leq 2\max_i E\left\|\frac{1}{T}\sum_{t=1}^{T} x_{it}^* u_{it}\right\|^2 + 2\max_i E\left\|\frac{1}{T}\sum_{1\leq t\neq s\leq T} x_{it}^* u_{is}\right\|^2.$$

It is easy to see that under Assumption A1(i)-(ii), the first term on the right hand side of the above equation is bounded from the above by

$$\max_i \frac{2}{T}\sum_{t=1}^{T} E\left\|x_{it}^* u_{it}\right\|^2 \leq 2\max_{i,t} E\left\|x_{it}^* u_{it}\right\|^2 \leq C.$$

For the second term, by straightforward moment calculations for second-order degenerate U-statitics (see, e.g., Lemma A.2(ii) in Gao (2007, p.194)), we have $2\max_i E\left\|\frac{1}{T}\sum_{1\le t\ne s\le T} x_{it}^* u_{is}\right\|^2 \le C$ under Assumption A.2*(i.a)-(i.b). Consequently, $\max_i E\left\|\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T x_{it}^* u_{is}\right\|^2 \le C$ for some $C<\infty$.

C.2 Verification of Assumptions A.4 and A.5

In this subsection, we verify the conditions in Assumption A.4(ii)-(iv) and A.5(i)-(iv) under some primitive conditions when $\{u_{it}, t\ge 1\}$ is a generic stationary and invertible ARMA process. For simplicity, we focus on the case where u_{it}'s are independent along the cross-section dimension.

The invertibility of the ARMA process implies that we can write $\{u_{it}, t\ge 1\}$ as an AR(∞) process and approximate it by an AR(p) process for sufficiently large p :

$$\begin{aligned} u_{it} &= \sum_{j=1}^{\infty}\rho_j u_{i,t-j} + e_{it} = \left(\sum_{j=1}^{p}\rho_j u_{i,t-j}\right) + \left(\sum_{j=p+1}^{\infty}\rho_j u_{i,t-j} + e_{it}\right) \\ &= \boldsymbol{\rho}'\underline{u}_{i,t-1} + v_{it}, \end{aligned}$$

where $\boldsymbol{\rho} = \left(\rho_1, ..., \rho_p\right)'$, $\underline{u}_{i,t-1} = (u_{i,t-1}, ..., u_{i,t-p})'$, $v_{it} = v_{it,p} + e_{it}$, $v_{it,p} = \sum_{j=p+1}^{\infty}\rho_j u_{i,t-j}$ signifies the approximation error, and e_{it} is the error term with mean zero and variance $\sigma_{i,e}^2$. Note that $v_{it,p} = 0$ if $\{u_{it}, t\ge 1\}$ is an autoregressive process of order p or less. Let $e_t = (e_{1t}, ..., e_{Nt})'$ and $x_t = (x_{1t}, ..., x_{Nt})'$. Let $\kappa_i\left(0, s_1, ..., s_{a-1}\right)$ denote the ath order joint cumulant of $\left(e_{i0}, e_{is_1}..., e_{is_{a-1}}\right)$ where $s_1, ..., s_{a-1}$, and a are integers. Let $\Phi\left(\cdot\right)$ be defined as in Section 2.3.

Assumption A.4* (i) $\sum_{j=1}^{\infty}\rho_j z^j \neq 0$ for any complex number z with $|z|\le 1$, $\sum_{j=1}^{\infty}\left|\rho_j\right| < \infty$, $p^{3/2}(N^{-1} + T^{-1}) = o\left(1\right)$, and $\left(NT\right)^{1/2}\sum_{j=p+1}^{\infty}\left|\rho_j\right| = O\left(1\right)$.

(ii) For each i, $\{e_{it}, t\ge 1\}$ is strictly stationary and ergodic such that $E\left(e_{it}|\mathcal{F}_{t-1}\right) = 0$ where $\mathcal{F}_{t-1} = \sigma\left(e_{t-1}, e_{t-2}, ..., x_t, x_{t-1}...\right)$ is the σ-field generated by $\{e_{t-1}, e_{t-2}, ..., x_t, x_{t-1}...\}$, $E\left(e_{it}^2\right) = \sigma_{i,e}^2$, and $\max_i \sum_{s_1=-\infty}^{\infty}\cdots\sum_{s_{a-1}=-\infty}^{\infty}\kappa_i\left(0, s_1, ..., s_{a-1}\right) < C$ for $a = 2, 3, 4$; $\{x_{it}, e_{it}, \alpha_i\}$ are independent along the individual dimension; $\frac{1}{N}\sum_{i=1}^{N}\sigma_{i,e}^2 \to \bar{\sigma}_e^2$ as $N\to\infty$.

(iii) $E\left(e_{it}\alpha_i\right) = 0$, $E\left(e_{it}|e_{t-1}, e_{t-2}, ..., \lambda_t, \lambda_{t-1}, ...\right) = 0$, $E(\alpha_i^4)\le C$, and $E(\lambda_t^4)\le C$.

Assumption A.4*(i) is similar to Assumption A.4(i) except that now we do not impose any condition on $\lambda_{\min}\left(\Gamma_p\right)$ but require $\left(NT\right)^{1/2}\sum_{j=p+1}^{\infty}\left|\rho_j\right| = O\left(1\right)$. Following Lee, Okui and Shintani (2018, LOS hereafter), we can easily show that the condition on $\lambda_{\min}\left(\Gamma_p\right)$ is satisfied under Assumption A.4*(i)-(ii). The condition $\left(NT\right)^{1/2}\sum_{j=p+1}^{\infty}\left|\rho_j\right| = O\left(1\right)$ is weaker than the requirement that $\left(NT\right)^{1/2}\sum_{j=p+1}^{\infty}\left|\rho_j\right| = o\left(1\right)$ in LOS because we do not consider bias correction in our setup. Assumption A.4*(ii) imposes that e_{it}'s are independent along the individual dimension and a martingale difference sequence (m.d.s.) along the time dimension. The independence assumption can be relaxed to allow for certain weak form cross-sectional dependence at more lengthy arguments. The m.d.s. sequence is also assumed in Gonçalves and Kilian (2007) in the time series setup and it is weaker than the i.i.d. requirement in Lewis and Reinsel (1985) and LOS. Note that Assumption A.4*(ii) implies that $\{u_{it}, t\ge 1\}$ is a stationary process for each i, and can be represented by an infinite order moving average (MA(∞)) process, and the approximation error

$v_{it,p} = \sum_{j=p+1}^{\infty} \rho_j u_{i,t-j}$ is well behaved in the sense of mean square errors. Assumption A.4*(iii) is used to verify A.4(iv).

First, we verify Assumption A.4(ii). Assumption A.4*(i)-(ii) ensures that $\{u_{it}\}$ has mean zero and finite fourth moment and $E(v_{it}) = \sum_{j=p+1}^{\infty} \rho_j E(u_{i,t-j}) + E(e_{it}) = 0$. Let $c_\rho = \sum_{j=p+1}^{\infty} |\rho_j|$. Then $v_{it,p} = 0$ if $c_\rho = 0$. Without loss of generality, we assume that $c_\rho > 0$. Note that Assumption A.4*(i) implies that $c_\rho = O((NT)^{-1/2}) = o(1)$ and $\sum_{j=p+1}^{\infty} |\rho_j|^4 \leq \max_{j \geq p+1} |\rho_j|^3 c_\rho \leq c_\rho^4 = o(1)$. Then by Jensen inequality

$$\begin{aligned}
\max_{i,t} E(v_{it,p}^4) &= \max_{i,t} E\left(\sum_{j=p+1}^{\infty} \rho_j u_{i,t-j}\right)^4 \leq c_\rho^4 \max_{i,t} E\left(\frac{1}{c_\rho}\sum_{j=p+1}^{\infty} |\rho_j| |u_{i,t-j}|\right)^4 \\
&\leq c_\rho^3 \max_{i,t} \sum_{j=p+1}^{\infty} |\rho_j|^4 E(u_{i,t-j}^4) \leq C c_\rho^3 \sum_{j=p+1}^{\infty} |\rho_j|^4 = o(1).
\end{aligned}$$

It follows that $\max_{i,t} E(v_{it}^4) \leq 8 \max_{i,t} E(v_{it,p}^4) + 8 \max_{i,t} E(e_{it}^4) \leq C < \infty$. In addition, by the law of large numbers, we have $\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} e_{it}^2 = \bar{\sigma}_e^2 + o_P(1)$, where $\bar{\sigma}_e^2 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_{i,e}^2$. So in this case, $\bar{\sigma}_v^2 = \bar{\sigma}_e^2$.

Next, we verify Assumption A.4(iii) for $\zeta_{it} = \breve{x}_{it}$ as the case for $\zeta_{it} = 1$ is easier. Noting that $\breve{x}_{it} = x_{it} - \underline{x}_{i,t-1}\boldsymbol{\rho}$, we have

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \breve{x}_{it} v_{it} = \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \breve{x}_{it} v_{it,p} + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \breve{x}_{it} e_{it} \equiv b_{1,1} + b_{1,2}, \text{ say.}$$

Note that

$$\begin{aligned}
E\|b_{1,1}\| &\leq \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} E\|\breve{x}_{it} v_{it,p}\| \leq \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \left\{E\|\breve{x}_{it}\|^2\right\}^{1/2} \left\{E\left\|v_{it,p}^2\right\|\right\}^{1/2} \\
&\leq C\sum_{j=p+1}^{\infty} |\rho_j| = O\left((NT)^{-1/2}\right)
\end{aligned}$$

where we use the fact that

$$\begin{aligned}
\max_{i,t} E\|\breve{x}_{it}\|^2 &= 2\max_{i,t} E\|x_{it}\|^2 + 2\max_{i,t} E\left\|\underline{x}_{i,t-1}\boldsymbol{\rho}\right\|^2 \leq 2\max_{i,t} E\|x_{it}\|^2 + 2\max_{i,t}\sum_{j=1}^{p}\sum_{j_1=1}^{p} \rho_j\rho_{j_1} E(x_{i,t-j}' x_{i,t-j_1}) \\
&\leq 2\max_{i,t} E\|x_{it}\|^2 \left\{1 + \left(\sum_{j=1}^{\infty} |\rho_j|\right)^2\right\} \leq C < \infty
\end{aligned}$$

and that by Assumption A.4*(i),

$$\max_{i,t} E(v_{it,p}^2) = \max_{i,t} \sum_{j=p+1}^{\infty}\sum_{j'=p+1}^{\infty} \rho_j \rho_{j'} E(u_{i,t-j} u_{i,t-j'}) \leq C\left(\sum_{j=p+1}^{\infty} |\rho_j|\right)^2 = O\left((NT)^{-1}\right).$$

It follows that $b_{1,1} = O_P\left((NT)^{-1/2}\right)$ by Markov inequality. For $b_{1,2}$, we have $E(b_{1,2}) = 0$ and for any nonrandom vector $\omega \in \mathbb{R}^k$ with $\|\omega\| = 1$,

$$\begin{aligned} \mathrm{Var}\left(\omega' b_{1,2}\right) &= \frac{1}{\left(NT_p\right)^2}\omega' \sum_{i=1}^{N} \mathrm{Var}\left(\sum_{t=p+1}^{T} \breve{x}_{it} e_{it}\right)\omega = \frac{1}{\left(NT_p\right)^2}\omega' \sum_{i=1}^{N}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T} E\left(\breve{x}_{it}\breve{x}_{is}' e_{it}e_{is}\right)\omega \\ &= \frac{1}{\left(NT_p\right)^2}\omega' \sum_{i=1}^{N}\sum_{t=p+1}^{T} E\left(\breve{x}_{it}\breve{x}_{it}' e_{it}^2\right)\omega = O((NT_p)^{-1}). \end{aligned}$$

Then $b_{1,2} = O_P\left((NT)^{-1/2}\right)$ by Chebyshev inequality. Consequently, we have shown that $\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} \breve{x}_{it} v_{it} = O_P\left((NT)^{-1/2}\right)$. Analogously, we can verify the last condition in A.4(iii).

Next, we verify Assumption A.4(iv). Note that

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}\alpha_i = \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it,p}\alpha_i + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} e_{it}\alpha_i \equiv b_{2,1} + b_{2,2}.$$

By Assumption A.4*(iii), we have $E\left(b_{2,2}\right) = 0$ and $\mathrm{Var}(b_{22}) = \frac{1}{(NT_p)^2}\sum_{i=1}^{N} E\left(\sum_{t=p+1}^{T} e_{it}\alpha_i\right)^2 = O\left(N^{-1}\right)$. So $b_{2,2} = o_P\left(1\right)$. $b_{2,1} = o_P\left(1\right)$ by Markov inequality and the fact that

$$E\left|b_{2,1}\right| \leq \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} E\left|v_{it,p}\alpha_i\right| \leq \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}\left\{E\left(\alpha_i^2\right)E(v_{it,p}^2)\right\}^{1/2} \leq C\sum_{j=p+1}^{\infty}\left|\rho_j\right| = o\left(1\right).$$

Similarly, we have

$$\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it}[\Phi\left(L\right)\lambda_t] = \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} v_{it,p}[\Phi\left(L\right)\lambda_t] + \frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T} e_{it}[\Phi\left(L\right)\lambda_t] \equiv b_{3,1} + b_{3,2}.$$

Following the analysis of $b_{2,1}$, we can readily show that $b_{3,1} = o_P\left(1\right)$ by Markov inequality. For $b_{3,2}$, we have $E\left(b_{3,2}\right) = 0$ and

$$\mathrm{Var}\left(b_{3,2}\right) = \frac{1}{N^2T_p^2}\sum_{i=1}^{N}\sum_{i_1=1}^{N}\sum_{t=p+1}^{T} E\left(e_{it}e_{i_1t}[\Phi\left(L\right)\lambda_t]^2\right) = O\left(T^{-1}\right).$$

So $b_{3,2} = o_P\left(1\right)$.

Next, we verify that Assumption A.5(i) is satisfied with $\bar{\sigma}_{v,1}^2 = \bar{\sigma}_e^2$. Noting that $\bar{v}_{i\cdot} = \frac{1}{T_p}\sum_{t=p+1}^{T}\left(v_{it,p} + e_{it}\right) = \bar{v}_{i\cdot,p} + \bar{e}_{i\cdot}$ with $\bar{v}_{i\cdot,p} = \frac{1}{T_p}\sum_{t=p+1}^{T} v_{it,p}$ and $\bar{e}_{i\cdot} = \frac{1}{T_p}\sum_{t=p+1}^{T} e_{it}$, we have

$$\frac{T_p}{N}\sum_{i=1}^{N}\left(\bar{v}_{i\cdot}\right)^2 = \frac{T_p}{N}\sum_{i=1}^{N}\left(\bar{v}_{i\cdot,p}\right)^2 + \frac{T_p}{N}\sum_{i=1}^{N}\bar{e}_{i\cdot}^2 + \frac{2T_p}{N}\sum_{i=1}^{N}\bar{v}_{i\cdot,p}\bar{e}_{i\cdot} \equiv a_{1,1} + a_{1,2} + a_{1,3}, \text{ say.}$$

$a_{1,1} = o_P\left(1\right)$ by Markov inequality and the fact that

$$\begin{aligned} E\left(a_{1,1}\right) &= \frac{T_p}{N}\sum_{i=1}^{N} E\left(\frac{1}{T_p}\sum_{t=p+1}^{T} v_{it,p}\right)^2 \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{t=p+1}^{T} E\left(v_{it,p}^2\right) \leq T\max_{i,t} E(v_{it,p}^2) \\ &\leq C\left(\sqrt{T}\sum_{j=p+1}^{\infty}\left|\rho_j\right|\right)^2 = o\left(1\right). \end{aligned}$$

By Assumption A.4*(ii), $E\left(a_{1,2}\right)=\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}E\left(e_{it}^2\right)=\frac{1}{NT_p}\sum_{i=1}^{N}\sigma_{i,e}^2=\bar{\sigma}_e^2+o(1)$ and

$$\operatorname{Var}\left(a_{1,2}\right)=\frac{T_p}{N^2}\operatorname{Var}\left\{\sum_{i=1}^{N}\left(\frac{1}{T_p}\sum_{t=p+1}^{T}e_{it}\right)^2\right\}=\frac{1}{N^2T_p}\sum_{i=1}^{N}\operatorname{Var}\left\{\left(\sum_{t=p+1}^{T}e_{it}\right)^2\right\}=O\left(N^{-1}\right).$$

It follows that $a_{1,2}=\bar{\sigma}_e^2+o_P(1)$. By the Cauchy-Schwarz inequality, $a_{1,3}\leq 2\left(a_{1,1}a_{1,2}\right)^{1/2}=o_P\left(1\right).$ Then $\frac{T_p}{N}\sum_{i=1}^{N}\left(\bar{v}_{i\cdot}\right)^2=\bar{\sigma}_e^2+o_P(1)$.

Next, we verify that Assumption A.5(ii) is satisfied with $\bar{\sigma}_{v,2}^2=\bar{\sigma}_e^2$. Noting that $\bar{v}_{\cdot t}=\frac{1}{N}\sum_{i=1}^{N}\left(v_{it,p}+e_{it}\right)=\bar{v}_{\cdot t,p}+\bar{e}_{\cdot t}$ with $\bar{v}_{\cdot t,p}=\frac{1}{N}\sum_{i=1}^{N}v_{it,p}$ and $\bar{e}_{\cdot t}=\frac{1}{N}\sum_{i=1}^{N}e_{it}$, we have

$$\frac{N}{T_p}\sum_{t=p+1}^{T}\left(\bar{v}_{\cdot t}\right)^2=\frac{N}{T_p}\sum_{t=p+1}^{T}\left(\bar{v}_{\cdot t,p}\right)^2+\frac{N}{T_p}\sum_{t=p+1}^{T}\bar{e}_{\cdot t}^2+\frac{2N}{T_p}\sum_{t=p+1}^{T}\bar{v}_{\cdot t,p}\bar{e}_{\cdot t}\equiv a_{2,1}+a_{2,2}+a_{2,3},\text{ say.}$$

$a_{2,1}=o_P\left(1\right)$ by Markov inequality and the fact that

$$\begin{aligned}E\left(a_{2,1}\right)&=\frac{N}{T_p}\sum_{t=p+1}^{T}E\left(\frac{1}{N}\sum_{i=1}^{N}v_{it,p}\right)^2\leq\frac{1}{T_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}E\left(v_{it,p}^2\right)\leq N\max_{i,t}E(v_{it,p}^2)\\&\leq C\left(\sqrt{N}\sum_{j=p+1}^{\infty}\left|\rho_j\right|\right)^2=o\left(1\right).\end{aligned}$$

By Assumption A.4*(ii), $E\left(a_{2,2}\right)=\frac{1}{NT_p}\sum_{i=1}^{N}\sum_{t=p+1}^{T}E\left(e_{it}^2\right)=\frac{1}{NT_p}\sum_{i=1}^{N}\sigma_{i,e}^2=\bar{\sigma}_e^2+o(1)$ and

$$\operatorname{Var}\left(a_{2,2}\right)=\frac{N}{T_p^2}\operatorname{Var}\left\{\sum_{t=p+1}^{T}\left(\frac{1}{N}\sum_{i=1}^{N}e_{it}\right)^2\right\}=\frac{1}{NT_p^2}\sum_{t=p+1}^{T}\operatorname{Var}\left\{\left(\sum_{i=1}^{N}e_{it}\right)^2\right\}=O\left(T^{-1}\right).$$

It follows that $a_{2,2}=\bar{\sigma}_e^2+o_P(1)$. By Cauchy-Schwarz inequality, $a_{2,3}\leq 2\left(a_{2,1}a_{2,2}\right)^{1/2}=o_P\left(1\right).$ Then $\frac{N}{T_p}\sum_{t=p+1}^{T}\left(\bar{v}_{\cdot t}\right)^2=\bar{\sigma}_e^2+o_P(1)$.

To verify Assumption A.5(iii), note that

$$\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{v}_{i\cdot}=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(x_{it}\right)v_{is}+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^*v_{is}\equiv a_{3,1}+a_{3,2},\text{ say.}$$

For $a_{3,1}$, we have

$$a_{3,1}=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(x_{it}\right)v_{is,p}+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(x_{it}\right)e_{is}\equiv a_{3,11}+a_{3,12},\text{ say.}$$

Under Assumption A.4*(i)-(ii), we can readily show that $a_{3,12}=O_P\left((NT)^{-1/2}\right)$ by Chebyshev inequality. For $a_{3,11}$, we have

$$E\left|a_{3,1}\right|\leq\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|E\left(x_{it}\right)\right\|\left|E\left(v_{is,p}\right)\right|\leq C\max_{i,s}\left[E\left\|v_{is,p}^2\right\|\right]^{1/2}=O((NT)^{-1/2}).$$

Then $a_{3,1}=O_P\left((NT)^{-1/2}\right)$. For $a_{3,2}$, we have

$$a_{3,2}=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^*v_{is,p}+\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}^*e_{is}\equiv a_{3,21}+a_{3,22},\text{ say.}$$

Following the verification of Assumption A.2(iii), we can readily show that $a_{3,22} = O_P\left(T^{-1}\right)$. For $a_{3,21}$, we have

$$E\left\|a_{3,21}\right\| \leq \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|E\left(x_{it}^{*}v_{is,p}\right)\right\| \leq C\max_{i,s}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E(v_{is,p}^{2})\right]^{1/2} \leq C\sum_{j=p+1}^{\infty}\left|\rho_j\right| = O((NT)^{-1/2}).$$

It follows that $a_{3,2} = O_P\left((NT)^{-1/2} + T^{-1}\right)$ and $\frac{1}{N}\sum_{i=1}^{N}\bar{x}_{i\cdot}\bar{v}_{i\cdot} = O_P\left((NT)^{-1/2} + T^{-1}\right)$.

To verify Assumption A.5(iv), note that

$$\frac{1}{T_p}\bar{x}_{\cdot t}\bar{v}_{\cdot t} = \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(x_{it}\right)v_{jt} + \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{it}^{*}v_{jt} \equiv a_{4,1} + a_{4,2}, \text{ say.}$$

For $a_{4,1}$, we have

$$a_{4,1} = \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(x_{it}\right)v_{js,p} + \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(x_{it}\right)e_{jt} \equiv a_{4,11} + a_{4,12}, \text{ say.}$$

Following the analysis of $a_{2,1}$, we can readily show that $a_{4,12} = O_P\left((NT)^{-1/2}\right)$ by Chebyshev inequality. For $a_{4,11}$, we have

$$E\left|a_{4,1}\right| \leq \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|E\left(x_{it}\right)\right\|\left|E\left(v_{jt,p}\right)\right| \leq C\max_{j,t}\left[E\left\|v_{jt,p}^{2}\right\|\right]^{1/2} = O((NT)^{-1/2}).$$

Then $a_{4,1} = O_P\left((NT)^{-1/2}\right)$. For $a_{4,2}$, we have

$$a_{4,2} = \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{it}^{*}v_{jt,p} + \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}x_{it}^{*}e_{jt} \equiv a_{4,21} + a_{4,22}, \text{ say.}$$

Following the verification of Assumption A.2(iv), we can readily show that $a_{4,22} = O_P\left(N^{-1}\right)$. For $a_{4,21}$, we have

$$\begin{aligned}
E\left\|a_{4,21}\right\| &\leq \frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left(x_{it}^{*}v_{jt,p}\right) \leq C\frac{1}{N^2T_p}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[E(v_{jt,p}^{2})\right]^{1/2} \\
&\leq C\sum_{j=p+1}^{\infty}\left|\rho_j\right| = O((NT)^{-1/2}).
\end{aligned}$$

It follows that $a_{4,2} = O_P\left((NT)^{-1/2} + T^{-1}\right)$ and $\frac{1}{T_p}\bar{x}_{\cdot t}\bar{v}_{\cdot t} = O_P\left((NT)^{-1/2} + N^{-1}\right)$.

D Choice of p in the Modified Jackknife

As discussed in Remark 9 in the main paper, there are several practical approaches to choose p in the modified jackknife method.

First, we can use a "rule of thumb" and let p increase with T, e.g., $p = \lfloor T^{1/4}\rfloor$, where $\lfloor T^{1/4}\rfloor$ is the nearest integer less than or equal to $T^{1/4}$.

Second, we can follow Lee, Okui, and Shintani (2018) by setting $p_{\max} = \lfloor T^{1/4} \rfloor$ and consider a general-to-specific testing procedure based on t-statistic until we reject the null. Specifically, we first run the following auxiliary regression using the pooled OLS

$$\hat{u}_{it}^{(4)} = \rho_1 \hat{u}_{i,t-1}^{(4)} + \rho_2 \hat{u}_{i,t-2}^{(4)} + ... + \rho_{p_{\max}} \hat{u}_{i,t-p_{\max}}^{(4)} + \tilde{v}_{it}$$

and test $\rho_{p_{\max}} = 0$ using t-statistics. If it is rejected, we conclude that $p = p_{\max}$. If we fail to reject it, we eliminate the $p_{\max}$th lag and run the regression with $p_{\max} - 1$ lags, and test $\rho_{p_{\max}-1} = 0$. We continue this procedure until we reject the null. Note that here $\hat{u}_{it}^{(4)}$'s are estimated. To take this into account, Wooldridge (2010, p. 311) argues that for the pooled OLS, we should use the fully robust standard errors (robust to both heteroskedasticity and serial correlation, see equation (7.26) in Wooldridge (2010, p. 171)). Another issue is that we need to choose the nominal level to decide whether to reject. In our simulations below, we choose the conventional 5% level.

Third, we can apply the information criteria, such as AIC and BIC, to the residuals obtained from Model 4 ($\hat{u}_{it}^{(4)}$) to determine p. For the implementation, see, e.g., Stock and Watson (2012, Section 14.5). In general, BIC gives a consistent estimator of p, and AIC tends to choose a relatively large p.

We conduct simulations to examine the finite sample performance of four methods above, labelled as rule of thumb, testing, AIC, and BIC, respectively. We consider three DGPs which are the same as those in Section 3.2 except that now the errors follow AR(1), MA(1) and ARMA(1, 1) processes, respectively. Specifically, u_{it} is generated respectively as

$$\begin{aligned}
&\text{DGP D.1:} \qquad u_{it} = 0.5u_{i,t-1} + e_{it}, \\
&\text{DGP D.2:} \qquad u_{it} = e_{it} + 0.5e_{i,t-1}, \text{ and} \\
&\text{DGP D.3:} \qquad u_{it} = 0.75u_{i,t-1} + e_{it} + 0.5e_{i,t-1},
\end{aligned}$$

where e_{it} is an $N(0, 1)$ random variable.

Tables D1-D3 present the simulations results for DGPs D.1-D.3, respectively. For DGPs D.1 and D.2 with AR(1) and MA(1) errors respectively, both CV* and CV** work well. For the ARMA errors, CV* works well when T is large and outperforms CV** in general. This suggests that CV** which is based on the Cochrane–Orcutt procedure relies on the AR(p) assumption more. Among the four methods of selecting p, there is no dominant one. When the sample size is large and CV* is used, all four methods can select the true model with a high probability.

Table D1: Frequency of the model selected with selected p (DGP D.1: $u_{it} = 0.5u_{i,t-1} + e_{it}$)

Selection of lag p in CV* and CV**	True M		Model 1				Model 2				Model 3				Model 4			
	Selected M	(N,T)	**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
	AIC	(10,10)	**.21**	.68	.01	.10	.01	**.87**	0	.12	0	.01	**.18**	.81	0	.02	0	**.98**
		(50,10)	**.04**	.87	0	.09	0	**.91**	0	.09	0	0	**.03**	.97	0	0	0	**1**
		(10,50)	**.27**	.72	0	0	0	**1**	0	0	0	0	**.22**	.78	0	0	0	**1**
		(50,50)	**.04**	.96	0	0	0	**1**	0	0	0	0	**.03**	.97	0	0	0	**1**
	BIC	(10,10)	**.89**	.11	0	0	.09	**.91**	0	0	.36	.04	**.49**	.10	.51	.12	.01	**.36**
		(50,10)	**1**	0	0	0	.20	**.78**	.02	0	.01	0	**1**	0	.43	0	.16	**.42**
		(10,50)	**.98**	.02	0	0	0	**1**	0	0	.39	.58	**.03**	0	.37	.62	0	**.01**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	BIC_2	(10,10)	**.08**	.67	.02	.23	0	**.74**	0	.25	0	.01	**.09**	.90	0	.01	0	**.99**
		(50,10)	**.01**	.87	0	.12	0	**.88**	0	.13	0	0	**.01**	.99	0	0	0	**1**
		(10,50)	**.22**	.77	0	.01	0	**.99**	0	.01	0	0	**.18**	.82	0	0	0	**1**
		(50,50)	**.06**	.94	0	0	0	**1**	0	0	0	0	**.04**	.96	0	0	0	**1**
	CV	(10,10)	**.25**	.69	.01	.05	.01	**.93**	0	.06	.01	.02	**.28**	.69	0	.03	.01	**.97**
		(50,10)	**.06**	.88	0	.06	0	**.94**	0	.06	0	0	**.06**	.94	0	0	0	**1**
		(10,50)	**.28**	.72	0	0	0	**1**	0	0	0	0	**.29**	.71	0	0	0	**1**
		(50,50)	**.04**	.96	0	0	0	**1**	0	0	0	0	**.05**	.95	0	0	0	**1**
rule of thumb $\lfloor T^{1/4} \rfloor$	CV*	(10,10)	**.58**	.36	.03	.03	.03	**.91**	0	.06	.01	.01	**.61**	.37	.01	.01	.03	**.95**
		(50,10)	**.69**	.26	.03	.01	0	**.95**	0	.05	0	0	**.74**	.27	0	0	0	**1**
		(10,50)	**.87**	.13	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
		(50,50)	**.99**	.01	0	0	0	**1**	0	0	0	0	**.99**	.01	0	0	0	**1**
	CV**	(10,10)	**.81**	.14	.04	.01	.42	**.51**	.05	.02	.01	.01	**.84**	.14	.01	.01	.58	**.40**
		(50,10)	**.95**	.01	.05	0	.31	**.44**	.23	.02	0	0	**.99**	.01	0	0	.69	**.31**
		(10,50)	**.93**	.07	0	0	.01	**.99**	0	0	0	0	**.95**	.06	0	0	.09	**.91**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
AIC	CV*	(10,10)	**.25**	.69	.01	.05	.01	**.93**	0	.06	.01	.02	**.28**	.69	0	.03	.01	**.96**
		(50,10)	**.06**	.88	0	.06	0	**.94**	0	.06	0	0	**.06**	.94	0	0	0	**1**
		(10,50)	**.90**	.10	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	CV**	(10,10)	**.25**	.69	.01	.05	.01	**.93**	0	.06	.01	.02	**.28**	.69	0	.03	.01	**.96**
		(50,10)	**.06**	.88	0	.06	0	**.94**	0	.06	0	0	**.06**	.94	0	0	0	**1**
		(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	.02	**.98**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
BIC	CV*	(10,10)	**.31**	.63	.02	.05	.01	**.92**	0	.06	.01	.02	**.34**	.64	.01	.02	.01	**.96**
		(50,10)	**.06**	.88	0	.06	0	**.94**	0	.06	0	0	**.06**	.94	0	0	0	**1**
		(10,50)	**.91**	.09	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	CV**	(10,10)	**.32**	.62	.02	.05	.09	**.84**	.01	.06	.01	.02	**.35**	.63	.01	.02	.11	**.87**
		(50,10)	**.06**	.88	0	.06	0	**.93**	0	.06	0	0	**.06**	.94	0	0	.01	**.99**
		(10,50)	**.95**	.05	0	0	0	**1**	0	0	0	0	**.96**	.04	0	0	.02	**.98**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
Testing	CV*	(10,10)	**.56**	.38	.03	.03	03	**.91**	0	.06	.01	0	**.58**	.40	.01	.01	03	**.95**
		(50,10)	**.69**	.26	.03	.01	0	**.95**	0	.05	0	0	**.74**	.27	0	0	0	**1**
		(10,50)	**.90**	10	0	0	0	**1**	0	0	0	0	**.89**	.11	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	CV**	(10,10)	**.74**	.21	.04	.02	.40	**.52**	.05	.03	.01	0	**.77**	.22	.01	.01	.55	**.43**
		(50,10)	**.95**	.01	.05	0	.31	**.44**	.23	.02	0	0	**.99**	.01	0	0	.69	**.31**
		(10,50)	**.94**	.06	0	0	0	**1**	0	0	0	0	**.95**	.05	0	0	.03	**.97**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

Table D2: Frequency of the model selected with selected p (DGP D.2: $u_{it} = e_{it} + 0.5e_{i,t-1}$)

Selection of lag p in CV* and CV**	True M		Model 1				Model 2				Model 3				Model 4			
	Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
		(N,T)																
	AIC	(10,10)	**.55**	.36	.04	.05	.01	**.88**	0	.11	.01	.01	**.53**	.45	.01	.01	.01	**.97**
		(50,10)	**.59**	.35	.02	.04	0	**.92**	0	.09	0	0	**.58**	.42	0	0	0	**1**
		(10,50)	**.65**	.35	0	0	0	**1**	0	0	0	0	**.57**	.43	0	0	0	**1**
		(50,50)	**.73**	.27	0	0	0	**1**	0	0	0	0	**.70**	.30	0	0	0	**1**
	BIC	(10,10)	**.99**	.01	0	0	.16	**.84**	0	0	.35	.01	**.63**	.01	.75	.05	.01	**.19**
		(50,10)	**1**	0	0	0	.46	**.44**	.11	0	0	0	**1**	0	.64	0	.29	**.07**
		(10,50)	**1**	0	0	0	0	**1**	0	0	.49	.43	**.08**	0	.47	.53	0	**0**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
	BIC_2	(10,10)	**.29**	.47	.06	.18	0	**.74**	0	.26	0	0	**.31**	.68	0	0	0	**1**
		(50,10)	**.41**	.49	.03	.07	0	**.88**	0	.12	0	0	**.42**	.58	0	0	0	**1**
		(10,50)	**.57**	.43	0	0	0	**.99**	0	.01	0	0	**.49**	.51	0	0	0	**1**
		(50,50)	**.78**	.22	0	0	0	**1**	0	0	0	0	**.75**	.25	0	0	0	**1**
	CV	(10,10)	**.62**	.33	.03	.02	.01	**.94**	0	.05	.01	.01	**.65**	.33	.02	.03	.01	**.94**
		(50,10)	**.68**	.27	.03	.02	0	**.94**	0	.06	0	0	**.72**	.28	0	0	0	**1**
		(10,50)	**.66**	.34	0	0	0	**1**	0	0	0	0	**.66**	.34	0	0	0	**1**
		(50,50)	**.75**	.25	0	0	0	**1**	0	0	0	0	**.75**	.25	0	0	0	**1**
rule of thumb $\lfloor T^{1/4} \rfloor$	CV*	(10,10)	**.82**	.12	.04	.01	.04	**.91**	0	.05	.01	0	**.86**	.13	.02	.01	.04	**.93**
		(50,10)	**.94**	.01	.05	0	0	**.94**	0	.06	0	0	**.99**	.01	0	0	0	**1**
		(10,50)	**.86**	.14	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
		(50,50)	**.99**	.01	0	0	0	**1**	0	0	0	0	**.99**	.01	0	0	0	**1**
	CV**	(10,10)	**.91**	.04	.05	0	.46	**.45**	.07	.02	.01	0	**.95**	.04	.02	.01	.67	**.31**
		(50,10)	**.94**	0	.06	0	.33	**.27**	.39	.01	0	0	**1**	0	0	0	.84	**.16**
		(10,50)	**.91**	.09	0	0	0	**1**	0	0	0	0	**.91**	.09	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
AIC	CV*	(10,10)	**.62**	.33	.03	.02	.01	**.94**	0	.05	.01	.01	**.65**	.33	.02	.03	.01	**.94**
		(50,10)	**.68**	.27	.03	.02	0	**.94**	0	.06	0	0	**.72**	.28	0	0	0	**1**
		(10,50)	**.92**	.08	0	0	0	**1**	0	0	0	0	**.92**	.08	0	0	0	**1**
		(50,50)	**.86**	.14	0	0	0	**1**	0	0	0	0	**.86**	.14	0	0	0	**1**
	CV**	(10,10)	**.62**	.33	.03	.02	.01	**.94**	0	.05	.01	.01	**.65**	.33	.02	.03	.01	**.94**
		(50,10)	**.68**	.27	.03	.02	0	**.94**	0	.06	0	0	**.72**	.28	0	0	0	**1**
		(10,50)	**.93**	.07	0	0	0	**1**	0	0	0	0	**.93**	.07	0	0	0	**1**
		(50,50)	**.86**	.14	0	0	0	**1**	0	0	0	0	**.86**	.14	0	0	0	**1**
BIC	CV*	(10,10)	**.62**	.33	.03	.02	.01	**.93**	0	.06	.01	.01	**.64**	.34	.02	.02	.01	**.95**
		(50,10)	**.68**	.27	.03	.02	0	**.94**	0	.06	0	0	**.72**	.28	0	0	0	**1**
		(10,50)	**.89**	.11	0	0	0	**1**	0	0	0	0	**.90**	.10	0	0	0	**1**
		(50,50)	**.99**	.01	0	0	0	**1**	0	0	0	0	**.99**	.01	0	0	0	**1**
	CV**	(10,10)	**.64**	.31	.04	.02	.05	**.88**	.01	.05	.01	.01	**.67**	.31	.02	.02	.08	**.88**
		(50,10)	**.68**	.27	.03	.02	0	**.94**	0	.06	0	0	**.72**	.28	0	0	0	**1**
		(10,50)	**.93**	.07	0	0	0	**1**	0	0	0	0	**.93**	.07	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**
Testing	CV*	(10,10)	**.80**	.15	.04	.01	.04	**.91**	0	.05	.01	0	**.84**	.15	.02	.01	.04	**.93**
		(50,10)	**.94**	.01	.05	0	0	**.94**	0	.06	0	0	**.99**	.01	0	0	0	**1**
		(10,50)	**.86**	.14	0	0	0	**1**	0	0	0	0	**.87**	.13	0	0	0	**1**
		(50,50)	**.99**	.01	0	0	0	**1**	0	0	0	0	**.99**	.01	0	0	0	**1**
	CV**	(10,10)	**.86**	.09	.05	.01	.43	**.49**	.06	.03	.01	0	**.91**	.08	.02	.01	.61	**.36**
		(50,10)	**.94**	0	.06	0	.33	**.27**	.39	.01	0	0	**1**	0	0	0	.84	**.16**
		(10,50)	**.91**	.09	0	0	0	**1**	0	0	0	0	**.91**	.09	0	0	0	**1**
		(50,50)	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0	**1**

Table D3: Frequency of the model selected with selected p (DGP D.3: $u_{it} = 0.75u_{i,t-1} + e_{it} + 0.5e_{i,t-1}$)

Selection of lag p in CV* and CV**	True M		Model 1				Model 2				Model 3				Model 4			
	Selected M		**M1**	M2	M3	M4	M1	**M2**	M3	M4	M1	M2	**M3**	M4	M1	M2	M3	**M4**
		(N,T)																
	AIC	(10,10)	**0**	.83	0	.16	0	**.84**	0	.16	0	.17	**0**	.82	0	.17	0	**.82**
		(50,10)	**0**	.87	0	.14	0	**.87**	0	.14	0	0	**0**	1	0	0	0	**1**
		(10,50)	**.01**	.95	0	.04	0	**.97**	0	.04	0	.32	**.01**	.67	0	.32	0	**.68**
		(50,50)	**0**	.97	0	.03	0	**.97**	0	.03	0	0	**0**	1	0	0	0	**1**
	BIC	(10,10)	**.14**	.85	0	.01	.06	**.93**	0	.01	.17	.66	**.01**	.17	.15	.68	0	**.17**
		(50,10)	**.16**	.84	0	0	.02	**.98**	0	0	.05	.05	**.10**	.80	.08	.05	0	**.87**
		(10,50)	**.37**	.63	0	0	.03	**.97**	0	0	.27	.73	**0**	0	.26	.74	0	**0**
		(50,50)	**.52**	.48	0	0	0	**1**	0	0	.20	.50	**.18**	.12	.27	.56	0	**.17**
	BIC_2	(10,10)	**0**	.74	0	.26	0	**.74**	0	.26	0	.07	**0**	.93	0	.07	0	**.93**
		(50,10)	**0**	.84	0	.17	0	**.84**	0	.17	0	0	**0**	1	0	0	0	**1**
		(10,50)	**.01**	.93	0	.06	0	**.94**	0	.07	0	.21	**.01**	.78	0	.21	0	**.79**
		(50,50)	**0**	.98	0	.02	0	**.98**	0	.02	0	0	**0**	1	0	0	0	**1**
	CV	(10,10)	**.01**	.88	0	.11	0	**.89**	0	.11	0	.27	**.01**	.72	.01	.27	0	**.72**
		(50,10)	**0**	.90	0	.11	0	**.90**	0	.11	0	0	**0**	1	0	0	0	**1**
		(10,50)	**.02**	.96	0	.02	0	**.98**	0	.02	0	.42	**.01**	.56	0	.43	0	**.57**
		(50,50)	**0**	.98	0	.02	0	**.98**	0	.02	0	0	**0**	1	0	0	0	**1**
rule of thumb $\lfloor T^{1/4} \rfloor$	CV*	(10,10)	**.37**	.56	.03	.05	.14	**.79**	.01	.06	.01	0	**.41**	.58	0	.01	.14	**.85**
		(50,10)	**.41**	.51	.03	.05	.01	**.90**	0	.08	0	0	**.46**	.54	0	0	.01	**.99**
		(10,50)	**.47**	.52	0	0	.02	**.97**	0	.01	0	0	**.48**	.52	0	0	.01	**.99**
		(50,50)	**.63**	.37	0	0	0	**.99**	0	.01	0	0	**.62**	.38	0	0	0	**1**
	CV**	(10,10)	**.76**	.19	.04	.02	.72	**.22**	.04	.02	0	0	**.81**	.19	0	0	.79	**.20**
		(50,10)	**.92**	.03	.05	0	.87	**.04**	.09	0	0	0	**.97**	.03	0	0	.97	**.03**
		(10,50)	**.82**	.18	0	0	.49	**.51**	0	0	0	0	**.82**	.18	0	0	.64	**.37**
		(50,50)	**.96**	.04	0	0	.46	**.54**	0	0	0	0	**.97**	.03	0	0	.71	**.29**
AIC	CV*	(10,10)	**.29**	.61	.02	.08	.10	**.79**	.01	.10	.01	.01	**.33**	.65	0	.02	.10	**.88**
		(50,10)	**.38**	.53	.03	.06	.01	**.89**	0	.09	0	0	**.42**	.58	0	0	.01	**.99**
		(10,50)	**.52**	.48	0	0	.03	**.96**	0	.01	0	0	**.52**	.48	0	0	.01	**.99**
		(50,50)	**.79**	.21	0	0	0	**1**	0	0	0	0	**.78**	.22	0	0	0	**1**
	CV**	(10,10)	**.62**	.30	.03	.05	.59	**.33**	.03	.05	0	.01	**.67**	.31	0	.01	.66	**.33**
		(50,10)	**.73**	.21	.04	.03	.69	**.21**	.07	.03	0	0	**.77**	.23	0	0	.77	**.23**
		(10,50)	**.85**	.15	0	0	.56	**.44**	0	0	0	0	**.85**	.15	0	0	.69	**.31**
		(50,50)	**.97**	.03	0	0	.73	**.27**	0	0	0	0	**.98**	.02	0	0	.84	**.16**
BIC	CV*	(10,10)	**.17**	.72	.02	.10	.05	**.83**	.01	.11	0	0	**.21**	.79	0	0	.06	**.94**
		(50,10)	**.04**	.85	0	.11	0	**.88**	0	.12	0	0	**.04**	.96	0	0	0	**1**
		(10,50)	**.48**	.52	0	0	.02	**.97**	0	.01	0	0	**.49**	.51	0	0	.01	**.99**
		(50,50)	**.63**	.37	0	0	0	**.99**	0	.01	0	0	**.62**	.38	0	0	0	**1**
	CV**	(10,10)	**.56**	.37	.03	.04	.50	**.43**	.03	.04	0	0	**.64**	.36	0	0	.60	**.40**
		(50,10)	**.54**	.40	.04	.03	.41	**.49**	.07	.03	0	0	**.59**	.41	0	0	.53	**.47**
		(10,50)	**.84**	.16	0	0	.53	**.47**	0	0	0	0	**.84**	.16	0	0	.67	**.33**
		(50,50)	**.96**	.04	0	0	.47	**.53**	0	0	0	0	**.97**	.03	0	0	.71	**.29**
Testing	CV*	(10,10)	**.37**	.56	.03	.05	.14	**.79**	.01	.06	.01	0	**.41**	.58	0	.01	.14	**.85**
		(50,10)	**.41**	.51	.03	.05	.01	**.90**	0	.08	0	0	**.46**	.54	0	0	.01	**.99**
		(10,50)	**.47**	.52	0	0	.02	**.97**	0	.01	0	0	**.48**	.52	0	0	.01	**.99**
		(50,50)	**.63**	.37	0	0	0	**.99**	0	.01	0	0	**.62**	.38	0	0	0	**1**
	CV**	(10,10)	**.76**	.19	.04	.02	.72	**.22**	.04	.02	0	0	**.81**	.19	0	0	.79	**.20**
		(50,10)	**.92**	.03	.05	0	.87	**.04**	.09	0	0	0	**.97**	.03	0	0	.97	**.03**
		(10,50)	**.82**	.18	0	0	.49	**.51**	0	0	0	0	**.82**	.18	0	0	.64	**.37**
		(50,50)	**.96**	.04	0	0	.46	**.54**	0	0	0	0	**.97**	.03	0	0	.71	**.29**

REFERENCES

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models, Econometrica 70, 191-221.

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

Gao, J. (2007). Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, New York.

Gonçalves, S., and L. Kilian (2007). Asymptotic and bootstrap inference for AR(∞) process with conditional heteroskedasticity, Econometric Reviews 26, 609-641.

Hahn, J. and G. Kuersteiner (2011). Reduction for dynamic nonlinear panel models with fixed effects, Econometric Theory 27, 1152-1191.

Lee, Y. J., R. Okui, and M. Shintani (2018). Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes, Journal of Econometrics, forthcoming.

Lewis, R., and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. Journal of Multivariate Analysis 16, 393-411.

Prakasa Rao, B. L. S. (2009). Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441-460.

Su, L. and Q. Chen (2013). Testing homogeneity in panel data models with interactive fixed effects, Econometric Theory 29, 1079-1135.

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.